STM CHARACTERIZATION OF INTERFACE ROUGHNESS

IN InAlAs / InGaAs SUPERLATTICES GROWN BY MOCVD

A Dissertation

by

FEDERICO LOPEZ CRUZ

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Michael B. Weimer
Committee Members,	Joseph H. Ross, Jr.
	Jairo Sinova
	Lawrence S. Brown
Head of Department,	George R. Welch

August 2015

Major Subject: Physics

ABSTRACT

We employ cross–sectional scanning tunneling microscopy to obtain an accurate statistical representation of the *as–grown* heterojunctions delineating the quantum wells and quantum barriers in InAlAs / InGaAs strain–balanced superlattices grown by MOCVD intimately related to quantum cascade lasers. Small deviations from the presumed planarity of these interfaces are believed to have profound implications for device performance, with energy–level broadening and carrier scattering as immediate, unintended consequences.

The aluminum–rich barrier layers in these structures present a significant challenge for present–day vacuum technology. Painstaking efforts are required to maintain a suitably–pristine habitat where freshly–exposed, aluminum–rich surfaces remain clean over the many days needed to conclude representative STM surveys.

We describe the development of carefully–constructed navigation protocols minimizing the image distortion inherent to STM piezo–electric raster mechanisms. These advances are used to implement a novel, reciprocal–space technique (analogous to Bragg's law in x–ray diffraction) that yields local distance metrics insensitive to STM raster non–idealities. The method's accuracy is demonstrated by way of local period measurements that agree (to within hundredths of monolayers) with high–resolution x-ray diffraction.

We develop robust image processing algorithms that incorporate statistical

criteria to reproducibly identify the interfaces separating quantum wells from barriers in cleavage–exposed cross–section. The heterojunction profiles obtained this way provide an experimentally–accessible avenue for delineating the confinement potential's spatial boundaries that appears logically consistent and physically reasonable.

We conduct a systematic analysis of interface roughness from the viewpoint of fluctuations about an experimentally–determined profile mean. Subtleties central to a correct understanding of roughness variance and its uncertainty in the face of correlated fluctuations are addressed, and the ensuing discussion illustrated with, and corroborated by, numerical simulations. Heterojunction growth order and growth–plane anisotropy are conclusively established as distinguishing physical characteristics.

We turn, finally, to the power spectra of these interface fluctuations to ascertain the correlation lengths and functional forms governing their respective spatial–frequency dependencies. The available data are consistent with a universal power spectrum for MOCVD roughness that is isotropic, independent of heterojunction growth order, and predominantly exponential in nature.

for my beautiful wife, daughter, son, and family

ACKNOWLEDGEMENTS

I would like to begin by recognizing the most important person responsible for my success at Texas A&M, my advisor Dr. Michael B. Weimer. He provided guidance and assistance that helped me accomplish the research project described in this manuscript. I am also grateful to his dedication as a mentor, and I sincerely believe that his influence will be responsible for many of my future goals.

I am also grateful for the assistance of the members of my committee: Dr. Joseph H. Ross, Jr., Dr. Jairo Sinova (who was also an essential link for my commencement as a graduate student at TAMU), and Dr. Lawrence S. Brown.

As part of a very collaborative research team, I am in extreme debt to (current and former) members of the STM–Laboratory: Kara M. Kanedy, who has always been of great support when data acquisition times exceeded reasonable working hours; Matthew R. Wood, who provided most of the computational machinery necessary to crack very complicated problems as well as our continuous and extensive discussions about semiconductors; Melissa Fuller, who laid ground on our preliminary image processing steps; Morgan A. Priolo, who helped us improve the sample cleaving protocol; Dr. Jeremy D. Steinshnider (a former member of the dynasty), who provided my first STM data acquisition lesson and extremely helpful technical insights.

Furthermore, I am grateful for the friendly crew of the Electronic shop, including Jason Caswell, who helped us pinpoint and solve most of our electronic obstacles; Steve Payne, for been able of managing such a helpful team; James Kirby, for his patience and

guidance in obtaining the correct electronic components and for becoming a mentor in my new biking hobby; Erwin Thomas, who ended up being a very good friend who provided exotic meats in times of food "scarcity." In addition, none of the exquisite data presented here would have been obtained without the assistance of the Machine shop members who helped us design and improve the vacuum chamber; Kavin Morris for his endless hours working with Autocad; Ronald Page (who I consider part of my family by now) for his talented and precise machining.

During my extensive stay at Texas A&M, I also became very good friends with people outside of the Physics Department, and I am very grateful to have met Dr. Raymundo Arroyave, whose encouragement throughout my last years was very important, as well as Alejandro Grados, who always encouraged me to remain optimistic even when things were discouraging.

I am also extremely thankful to Dr. Sergio Y. Rodriguez, whom I have known since studying for my Masters Degree at the University of Texas at El Paso and consider a very good friend of my family. He was always a great partner while studying for finals and always available to discuss any subject.

However, it is to my family that I am most grateful: My brother, Dr. Nazario Lopez–Cruz, whose knowledge kept me focused when I most needed it. His support has been an invaluable part of my achievements; my sister, Margarita Lopez–Cruz, who has always been very close to me and values spending time talking about the most random topics. Special appreciation goes to my parents, Miguel Lopez and Teresa Cruz, for their never–ending guidance that shaped me into the person I am today, and all of their

sacrifices that have allowed my siblings and I to always have the best possible education. I know you wanted one of us to follow in your footsteps as a medical doctor, and now you have two, albeit doctors in philosophy. You did a great job as parents.

Finally, I would like to thank the Lopez–Romo clan, my beautiful wife Idania, who has always been the support I needed most. I am extremely grateful for her patience, and always believing in me; my stellar daughter Isabella, who is always thrilled to know anything and everything about nature; and my son, Elias, for always finding a way to break the sturdiest objects. This work is for you and your future. I hope I'll be able to make–up for all the times you had to wait in–line behind the computer until I could come and play with you.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Technological Motivation

Quantum Cascade Lasers (QCL) are widely used for molecular trace gas identification and analysis in areas such as health, safety, and the environment. QCLs also have the potential to be used for free space telecommunications where typical atmospheric gasses offer broad wavelength ranges with minimal molecular absorption. The wide variety of QCL applications stems from its principle of operation. In 1971, Kazarinov *et. al*., [1] proposed a novel idea to employ an electrically biased superlattice [2] (SL) to achieve optical gain (laser emission) in the infrared (IR) spectrum, with the use of electron intersubband transitions solely between energy states in the conduction band. This innovative idea differs from the traditional semiconductor diode laser [3, 4] principle whereby optical emission occurs by electron–hole pair recombination from the conduction to valence band across the semiconductor band gap, and therefore enslaves the wavelength emission to the choice of the host semiconductor material. In 1994, Faist *et. al*. [5] demonstrated experimentally a semiconductor injection laser that achieved optical emission relying only on intersubband electron transitions between conduction band states.

QCLs are made of alternating layers of semiconductor materials (typically III – Vs), that can be tailored compositionally and structurally on the sub–nanometer scale to

emit a specific wavelength in the 3–14 μm (mid–IR) spectrum, a spectral region with few suitable radiation sources [6]. Moreover, in an appropriately biased QCL a single electron can be the predecessor of multiple emitted photons, each of them produced in the coupled "staircase" design of multiple quantum wells (MQW) delineated by the conduction band offset of the heterostructure as shown in Fig. 1.1. This ingenious unipolar design eliminates the need to simultaneously engineer both electron and hole mobilities.

Since the emission wavelength is determined by quantum confinement in the conduction band, and is therefore independent of the bulk band gap, QCLs can be fabricated with technologically robust, relatively–wide–gap semiconductor materials such as indium–phosphide [5], gallium–arsenide [7], and gallium–antimonide [8]. These materials have a long history of being successfully used for high–performance devices in the optoelectronics industry. This is both a financial bonus on the technology side that greatly reduces production costs, as well as scientifically advantageous since the detailed materials knowledge acquired during the optoelectronics era can (in principle) be adapted to the fabrication of QC devices.

That the vibrational modes of environmentally and medically significant small molecules lie in the infrared portion of the electromagnetic spectrum is a major driving force behind the rapid development of QC devices. Such heteronuclear molecules typically possess a characteristic vibrational mode in the mid–IR that uniquely identifies them and is often used for their accurate quantification. Simplified examples of typical vibrational modes exhibited in selected heteronuclear molecules are shown in Fig. 1.2 .

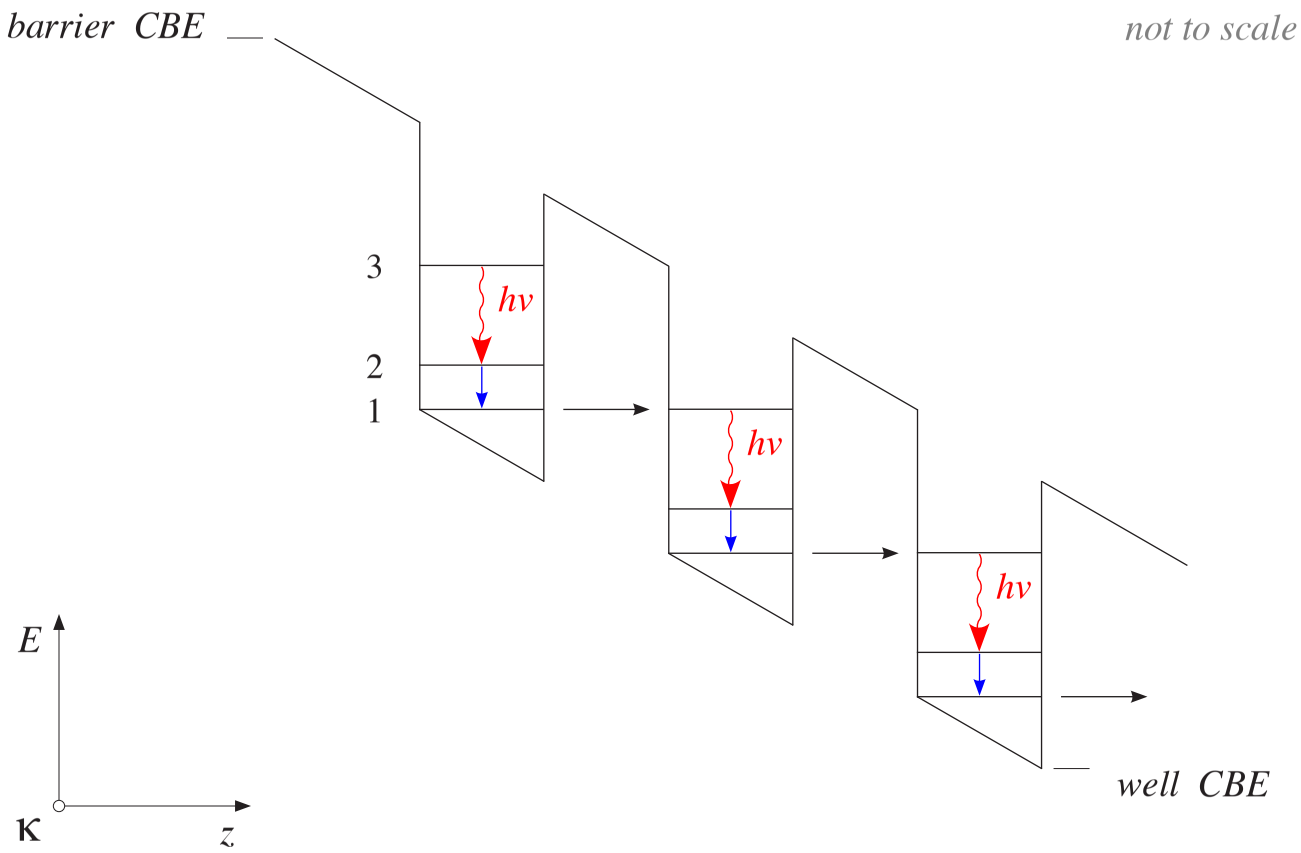


FIGURE 1.1. Conduction band profile for a simplified three–level quantum cascade laser under operating bias. Intersubband energy levels (1,2 and 3) are shown as horizontal lines within each well. The electron energy as a function of position assumes a staircase shape under the influence of an external electric field. Red arrows indicate the lasing transition and blue arrows the optical–phonon assisted depopulation of the lasing ground state in each "active region". Black arrows represent electrons "injected" into subsequent stages by tunneling through the barrier that separates adjacent QWs

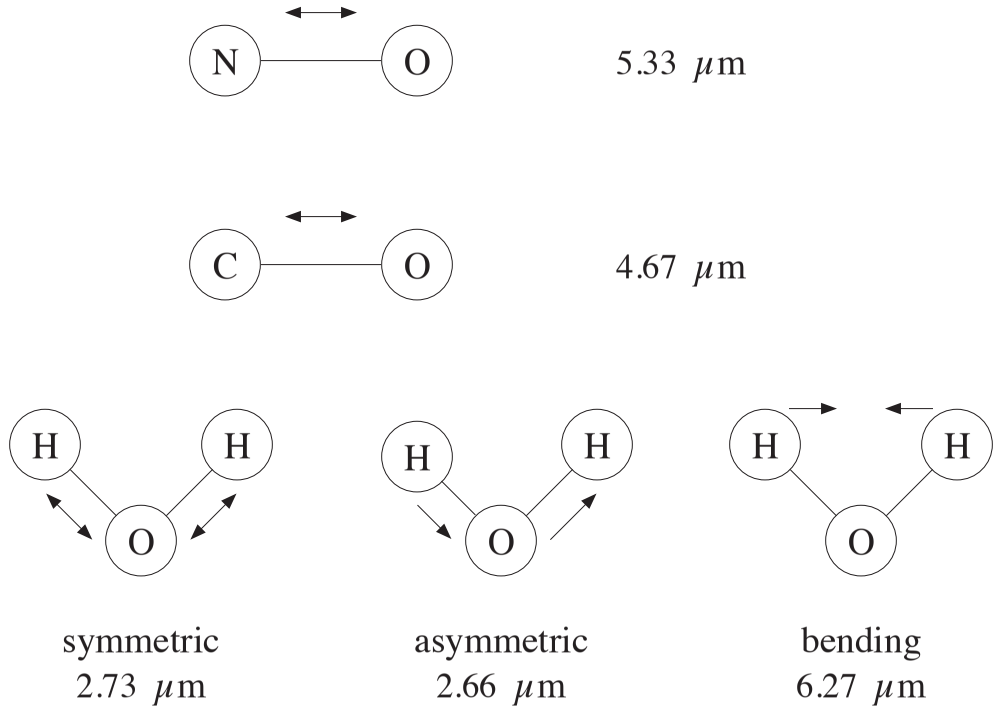


FIGURE 1.2. Illustration of vibrational modes present in IR active heteronuclear molecules. IR active molecules absorb electromagnetic radiation via a change in dipole moment accompanied by a variation in the internuclear separation. Linear molecules possess 3N–5 (non–linear molecues 3N–6) vibrational modes, where N is the number of atoms. Water, for example, possesses three vibrational modes illustrated here.

Chromatography and mass spectrometry systems are presently used for similar purposes by the industrial and scientific communities, however these instruments are large in size, complex to operate, easy to break, and relatively expensive. QCLs, on the other hand, are compact, simple, robust, economical devices in which the optical detection mechanism requires minimal sample material pre–treatment, is non–invasive, and is very fast.

Quantum Cascade Laser Limitations

Direct investigation of material properties that have immediate influence on QC device performance is understandably of scientific as well as technological importance. QCL deficiencies including carrier scattering and energy–level broadening are associated with interface roughness [9] and layer thickness non–uniformity, physical properties that, as we will see, are in principle observable with cross–sectional scanning tunneling microscopy (STM). Such deviations from the intended design have serious consequences for device performance; it is therefore important to understand these physical processes in sufficient detail that growth protocols favoring the formation of high–quality heterostructures may be adopted.

The physical connection between interface fluctuations and carrier scattering in MQWs will be addressed next, but we begin with a more general approach to quantum–mechanical scattering based on Fermi's golden rule [10]

$$\Gamma_{if} = \frac{2\pi}{\hbar} \left| \left\langle \psi_f \right| H_{pert}(\mathbf{r}) \left| \psi_i \right\rangle \right|^2 \delta \left(E_f - E_i \right). \qquad (1.1)$$

This formula connects the transition probability per unit time – or transition rate – between specified initial and final states of the unperturbed problem to a given, time–independent perturbation, H_{pert} (**r**). The δ function ensures energy–conserving transitions, since a time–independent perturbation cannot supply or remove energy from the system.

For the particular but important instance of elastic plane–wave scattering, the transition rate in (1.1) is proportional to

$$\left| \left\langle e^{i\,\mathbf{k}'\cdot\mathbf{r}} \right| H_{pert}(\mathbf{r}) \left| e^{i\,\mathbf{k}\cdot\mathbf{r}} \right\rangle \right|^2, \qquad (1.2)$$

where **r** and **k** are three–dimensional (3–D) vectors in real and reciprocal space, respectively. Equation (1.2) is the power spectrum of the perturbation, H_{pert} (**r**), with respect to the 3–D momentum transfer, **q** = **k**' – **k**. For a perturbation of known geometry, this power spectrum (and hence the scattering rate) is usually calculated directly.

It is useful to recall that power spectra and autocorrelation functions quite generally form reciprocal Fourier–transform pairs [11]. This fact is not emphasized in standard quantum–mechanical treatments where the geometry of the scattering potential is known and the ACF then trivial and irrelevant. A complete statistical description of

any (presumed) stationary stochastic [11] perturbation, such as interface roughness, is in principle supplied by either the (direct–space) autocorrelation function or its corresponding (reciprocal–space) power spectral density.

For the case of carrier scattering in MQWs, where the confining potential breaks three–dimensional translational symmetry, there is a two–dimensional analog to (1.2) that may be obtained from the 2–D Fourier transform – or power spectrum – of the 2–D perturbation potential. Consider an ideal quantum well (perfectly planar interfaces) where separation of variables dictates bound–state wave functions in the growth direction, z [1], multiplied by freely–propagating, plane–wave solutions in the perpendicular *x, y* plane. Now imagine an additional perturbation, $H_{pert}(\boldsymbol{\rho})$, that depends only on the two–dimensional vector $\boldsymbol{\rho}$ lying within a given heterojunction plane. Omitting all dependence on *z*, (1.2) then takes the form

$$\left| \left\langle e^{i\,\boldsymbol{\kappa}'\cdot\boldsymbol{\rho}} \right| H_{pert}(\boldsymbol{\rho}) \left| e^{i\,\boldsymbol{\kappa}\cdot\boldsymbol{\rho}} \right\rangle \right|^2 , \tag{1.3}$$

where $\boldsymbol{\kappa}$ is now a vector in the two–dimensional space reciprocal to $\boldsymbol{\rho}$. This formulation offers a general framework for thinking about 2–D plane–wave scattering irrespective of the perturbation potential's detailed physical origin. Any combination of factors, such as

[1] This symbol (*z*) will be re–used in several other, context–dependent circumstances throughout this manuscript: for example, in Chapter II, we will refer to the out–of–plane motion of the *z*–piezo that controls the vertical elevation of an STM tip with respect to a cleavage surface; in Chapter III we will use *z* to describe heterojunction height profiles, abandoning the less intuitive nomenclature introduced in (1.6) below, to unambiguously distinguish coordinates from surfaces that are functions of these coordinates.

monolayer fluctuations in barrier or well widths, alloy clustering, and point impurities – either alone or in conjunction – may contribute to the deviations from perfect translation invariance incorporated into $H_{pert}(\boldsymbol{\rho})$. In all such cases, however, the natural way to parameterize scattering from the resulting stochastic potential, $H_{pert}(\boldsymbol{\rho})$, is via the 2–D power spectral density (with respect to $\boldsymbol{\kappa}'$–$\boldsymbol{\kappa}$) that appears in (1.3).

A well–established model [12] from the literature offers a concrete illustration. Assume, again, an infinite square well with perfectly planar interfaces. The quantized–energy levels are given by

$$E_n = \frac{\pi^2 \hbar^2}{2m^* L_w^2}\, n^2 , \tag{1.4}$$

where m^* is the effective mass of electrons in the well, n is the quantization number, and L_w the width of the well in the growth direction. Now consider a *small* change in well width caused by a monolayer–like step in the 2–D interface profile, $h(\boldsymbol{\rho})$; this step introduces a local perturbation to the quantized–energy levels, $\delta E(\boldsymbol{\rho})$, given by

$$\delta E(\boldsymbol{\rho}) \;=\; \frac{\delta E_n}{\delta L_w}\, h(\boldsymbol{\rho}) \;=\; -\frac{\pi^2 \hbar^2}{m^* L_w^3}\, n^2\, h(\boldsymbol{\rho}) . \tag{1.5}$$

Replacing $H_{pert}(\boldsymbol{\rho})$ in (1.3) by $\delta E(\boldsymbol{\rho})$ from (1.5), we see this perturbation potential is *linear* in the *fluctuations* of the interface *profile* about its mean value. Again omitting all

terms not explicitly dependent on $\boldsymbol{\rho}$, the two–dimensional, roughness–induced scattering rate is then proportional to

$$\left| \left\langle e^{i\,\boldsymbol{\kappa}'\cdot\boldsymbol{\rho}} \right| h(\boldsymbol{\rho}) \left| e^{i\,\boldsymbol{\kappa}\cdot\boldsymbol{\rho}} \right\rangle \right|^2 . \tag{1.6}$$

This is just the power spectral density (as a function of the in–plane momentum transfer, $\boldsymbol{\kappa}'–\boldsymbol{\kappa}$) for the fluctuating interface profile, $h(\boldsymbol{\rho})$, conveniently referred to as a *terrace height* function.

It is often assumed [12 – 14], both on the basis of mathematical simplicity as well as the wide variety of physical systems exhibiting similar behavior, that the two–dimensional power spectral density is governed by a single–length–scale, isotropic Gaussian of the form

$$2\pi\ \Delta^2\ \Lambda^2\ \exp\left(-2\pi^2\ \Lambda^2\ q^2\right). \tag{1.7}$$

Here Δ is the roughness amplitude (Δ^2 the variance), Λ the Gaussian in–plane correlation length, and $q = \boldsymbol{\kappa}'–\boldsymbol{\kappa}$ the 2–D momentum transfer. Though the mathematics of this model appears perfectly general, its physical limitations are obvious, since it incorporates only one of the several possible sources of interface roughness described above.

Characterization of the fluctuating *confinement potential* in (1.6) will depend on experimentally–defined heterojunction profiles, $h_{expt}(\boldsymbol{\rho})$ (Chapter III), from which one might hope to obtain

$$\left| \left\langle e^{i\,\boldsymbol{\kappa}'\cdot\boldsymbol{\rho}} \right| h_{expt}(\boldsymbol{\rho}) \left| e^{i\,\boldsymbol{\kappa}\cdot\boldsymbol{\rho}} \right\rangle \right|^2, \tag{1.8}$$

but this is not realistic, given a full, two–dimensional mapping of interface fluctuations is beyond experimental reach. At best, one may presume – under an assumption of statistical independence with $h(x,y) = h_x(x) \bullet h_y(y)$ – that (1.7) can reasonably be replaced by

$$2\pi\,\Delta^2 \cdot \Lambda_x \exp\left(-2\pi^2\,\Lambda_x^2\,q_x^2\right) \cdot \Lambda_y \exp\left(-2\pi^2\,\Lambda_y^2\,q_y^2\right), \tag{1.9}$$

where x, and y indices here refer to the two independent (orthogonal) growth–plane directions (i.e. [110] and [1–10]) revealed in cleavage cross section.

Our study will attempt to address the following fundamental questions concerning epitaxial growth: first, is the observed roughness correlation actually a single–length–scale Gaussian despite the multiple physical mechanisms potentially contributing to heterojunction fluctuations? Second, are the two sides of each QW equivalent in actual devices, as tacitly assumed in (1.9)? Reflection symmetry will be broken if, as we have reason to suspect, InAlAs–on–InGaAs and InGaAs–on–InAlAs

heterojunctions have distinguishable roughness properties.

In order to systematically approach these questions we assemble statistically representative heterojunction ensembles from MOCVD–grown, strain–balanced superlattices (Chapter III), calculate their respective roughness amplitudes (Chapter IV), and power spectral densities (Chapter V), interpret the results physically, and briefly consider the implication of these findings for carrier scattering in working QCLs (Chapter VI).

Epitaxial Deposition Techniques

Epitaxial growth of the stacked single–crystal semiconductor materials required to form MQWs in QCLs takes advantage of two distinct technologically mature methods such as solid–source molecular beam epitaxy (MBE) [15] or gas–source metal–organic chemical vapor deposition (MOCVD) [16]. Both technologies have proven to be especially reliable in producing state–of–the–art MQWs.

In a typical MBE system elemental sources – for example indium, arsenide, gallium, and aluminum – are individually heated in effusion cells installed in an ultra–high vacuum (UHV) system. Computer–controlled pneumatic shutters, as indicated in Fig. 1.3, are sequentially opened, or closed, to direct elemental beams in vapor form towards a single (suitably heated) substrate that is continuously rotated to facilitate uniform growth. Due to the extremely precise control over epitaxial growth achievable with MBE it is universally considered the preferred growth technology for the development of semiconductor devices that are challenging to fabricate. This precision,

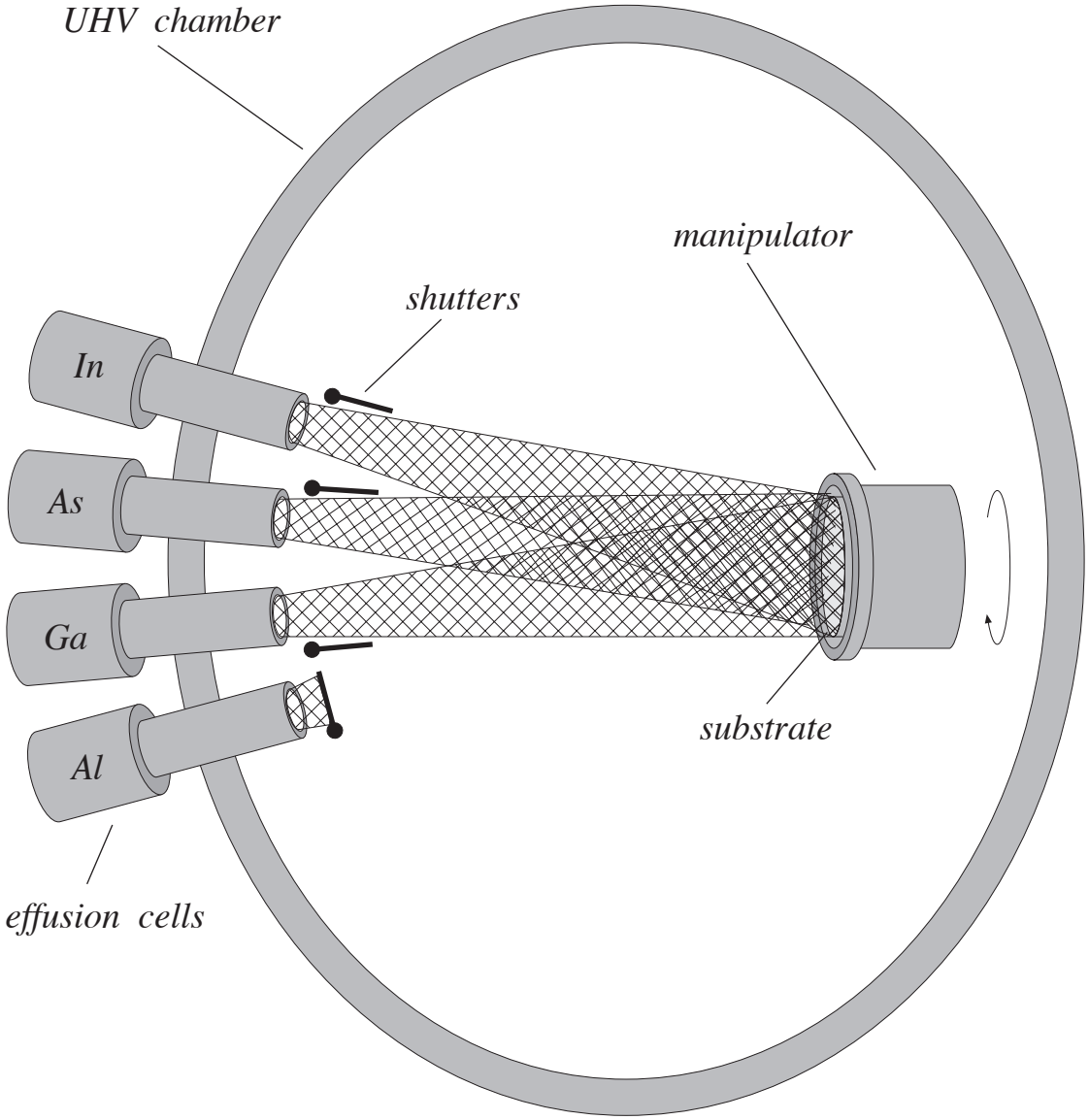


FIGURE 1.3. Simplified schematic of a typical MBE system enclosed in an ultra–high–vacuum chamber. Shutter arrangement shown here illustrates the growth of InGaAs. Substrate is continuously rotated to improve layer uniformity.

however, does not extend across the entire wafer as the geometry of the effusion cells produces layer thickness non–uniformity in the growth plane. MBE systems also suffer from prolonged down times due to (either planned or unplanned) maintenance, which is common of any apparatus residing in UHV environments.

MOCVD, on the other hand, is a more industry friendly, cost effective epitaxial growth technique. This technology uses more complex compound sources, for example trimethyl–indium, –gallium, –aluminum, arsine, and phosphine that are transported in precise dosages by a carrier gas flow at moderate pressures (10 to 760 Torr). The stream of selected metalorganic molecules is gently directed towards a heated substrate where a pyrolytic thermochemical reaction occurs resulting in epitaxial deposition of the metallic atoms on the substrate surface. Essential components of a standard vertical MOCVD setup capable of simultaneously growing multiple wafers are indicated in Fig. 1.4. Since these reactors operate at reasonable pressures prolonged downtime due to maintenance is far less frequent than with MBE systems.

The absence of high–vacuum in MOCVD reactors limits the number of *in–situ* monitoring techniques cable of providing immediate feedback during growths. Another technical inconvenience is the level of difficulty posed by growth interruptions, which are easily employed in typical MBE systems to improve the quality of interfaces. Finally, an important regulatory burden associated with all MOCVD systems is the need for expensive safety precautions to mitigate any environmental impact of the toxic gasses used during growth.

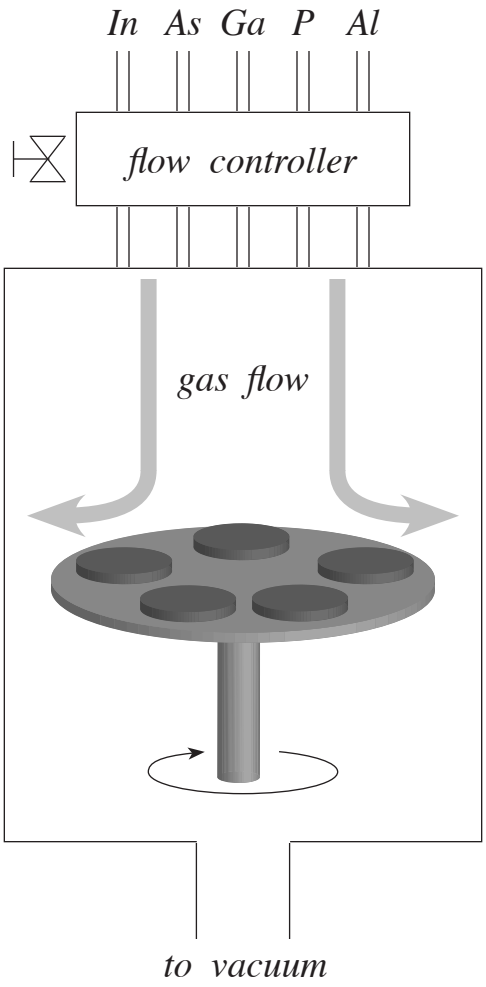


FIGURE 1.4. Simplified schematic of a vertical MOCVD reactor. Trimethyl precursors are used to transport cation elements while anions are transported by hydrides as arsine and phosphine. Multiple semiconductor wafers can be grown in a single run.

While inferior in terms of *in–situ* monitoring, growth control, and environmental safety, MOCVD still offers a competitive alternative to MBE and excels as the industry oriented growth method capable of satisfying the high–demand and high–performance mass–production inherently necessary for large–scale deployment of QCL technologies.

The detailed mechanism underlying epitaxial growth via MOCVD is complex and for many years the entire process was regarded as a "black box" [17] for which the standard measure of success was device performance. Additional details concerning MOCVD can be found elsewhere [16] and will not be addressed further, here.

Characterization Techniques Applicable to MOCVD

Real–time, layer–by–layer assessment of material quality during MOCVD growth is particularly challenging, since its relatively high–pressure environment does not support electron based methodologies, such as reflection high–energy electron diffraction (RHEED), routinely used for such monitoring during MBE. Options, here, are limited to optical techniques such as reflectance difference spectroscopy (RDS) [18, 19]. Developed primarily as a research tool during the 1990s, RDS is rarely, if ever, implemented on an industrial production scale. As a practical matter then, MOCVD growers must turn to high–resolution x–ray diffraction (HRXRD) and, to a lesser extent, to transmission electron microscopy (TEM) for any insights concerning structural quality, and photoluminescence (PL) for any insights concerning compositional control that bear on the bound–state spectrum of QW heterostructures.

Regardless of whether the assessment is made in real time or after the fact, these characterization tools (excepting TEM) share a common weakness: the material properties they reflect are averages over macroscopic length scales (typically microns to hundreds of microns), oftentimes in both lateral and vertical dimensions. Thus, meaningful structural and compositional insights on the nanometer scales relevant to electron confinement are difficult, if not impossible, to obtain, as are relevant details concerning lateral variability within the growth plane or layer–by–layer variability in the vertical growth direction.

Cross–sectional scanning tunneling microscopy (XSTM) is capable of analyzing QW heterostructures with extraordinary detail from growth initiation to completion through real–space characterization on the atomic–scale. Indeed, the technique has been successfully applied to a variety of III–V QW heterostructures grown by MBE [20 – 23]. However, the full potential of XSTM to enhance our understanding of QC structures has not yet been realized with either MBE– or MOCVD–grown material. The STM analyses of QCLs previously reported in the literature [24, 25] were not capable of quantitatively addressing the core material non–idealities inherent to these MQW structures – including layer thickness non–uniformity and interface fluctuations – that so profoundly influence device behavior. Here we propose to do exactly that.

Quantum Cascade Active Region versus Superlattice Test Structure

The active region of a working device (i.e., the region where photon emission occurs) is most often comprised of InGaAs quantum wells, and InAlAs barriers, lattice

matched to InP. Fig. 1.5 depicts approximately four QC active region repeats (grown, in this case, by MBE) obtained with cross–sectional TEM [26]. Active regions are typically designed with complex, *aperiodic* sequences that alternate comparatively thick wells (~ 17 ML) and very thin barriers (~ 4 ML) as shown schematically in Fig. 1.6, left. Accurate characterization of such a spatially non–uniform structure is a formidable task.

To simplify the experimental problem we have chosen to work with a *periodic* structure (an A–B–A–B … superlattice) in which all of the wells and barriers remain in the same proportion (see Fig. 1.6, right). The constituent layers thicknesses in this case are designed to mitigate the influence any one interface might exert on the quality of subsequently deposited ones. One may thus hope to obtain a representative analysis of the heterojunction statistical properties affecting carrier scattering that is independent of the particular set of layers over which STM data is actually acquired.

Scanning Tunneling Microscopy

STM Provides unprecedented structural detail concerning the epitaxial layers exposed in cross section by cleavage in vacuum. The exquisite atomic detail potentially achievable with STM is a direct consequence of the exponential dependence of the tunneling current on the spatial tip–sample separation. A one–dimensional view of this technique appears in Fig. 1.7 for an idealized, planar tip–sample junction where a negative bias voltage applied to the sample is referenced to the Fermi level of the tip [27]. The tunneling current is proportional to

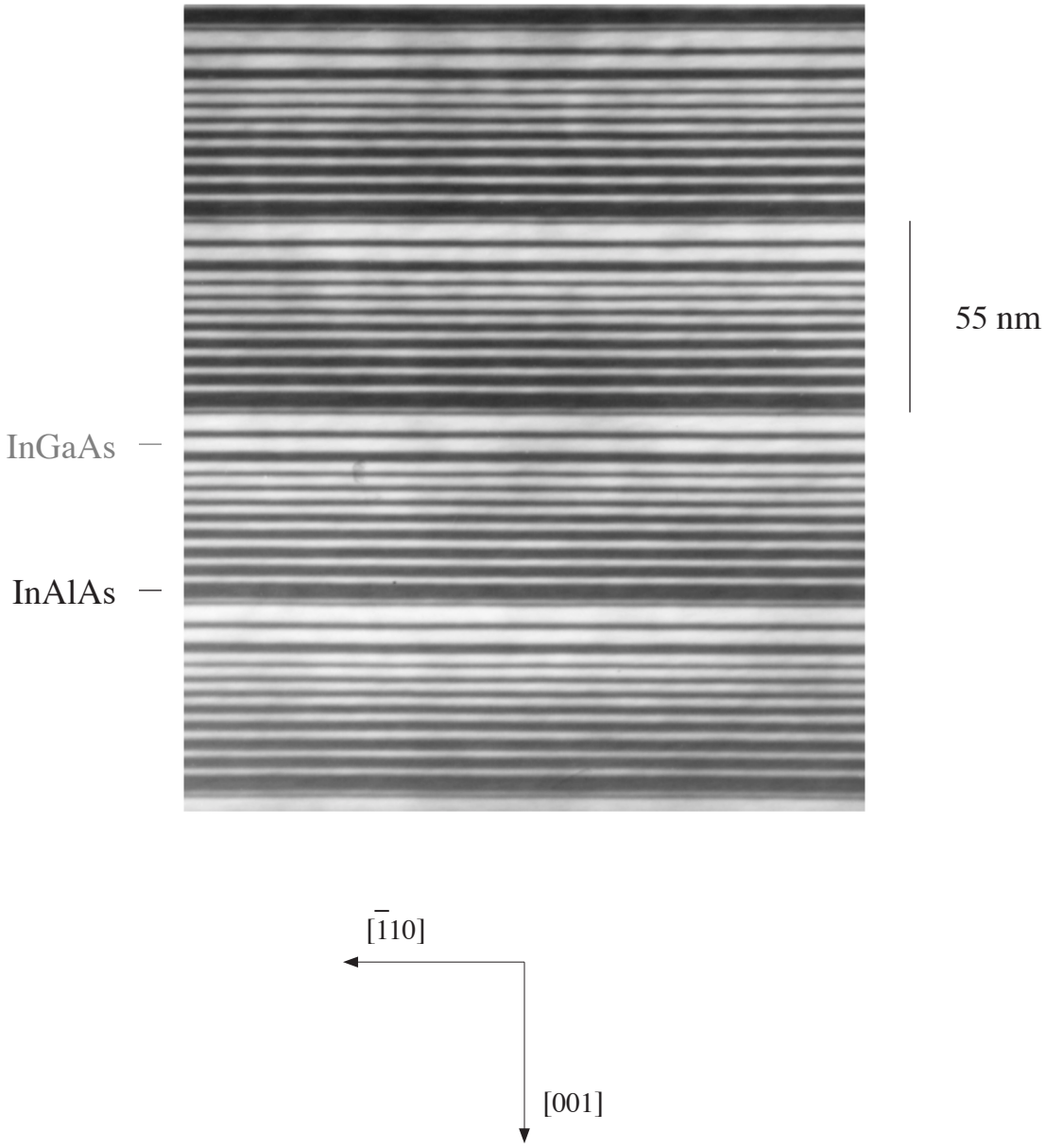


FIGURE 1.5. Cross–sectional TEM image showing approximately four QCL stages grown by MBE. Layer contrast is reversed here to facilitate comparison with STM. Each stage consists of an aperiodic sequence of several comparatively thick quantum wells and very thin quantum barriers. The complex layer geometry is clearly evident in this image. Growth direction is top to bottom. Reprinted with permission from [26].

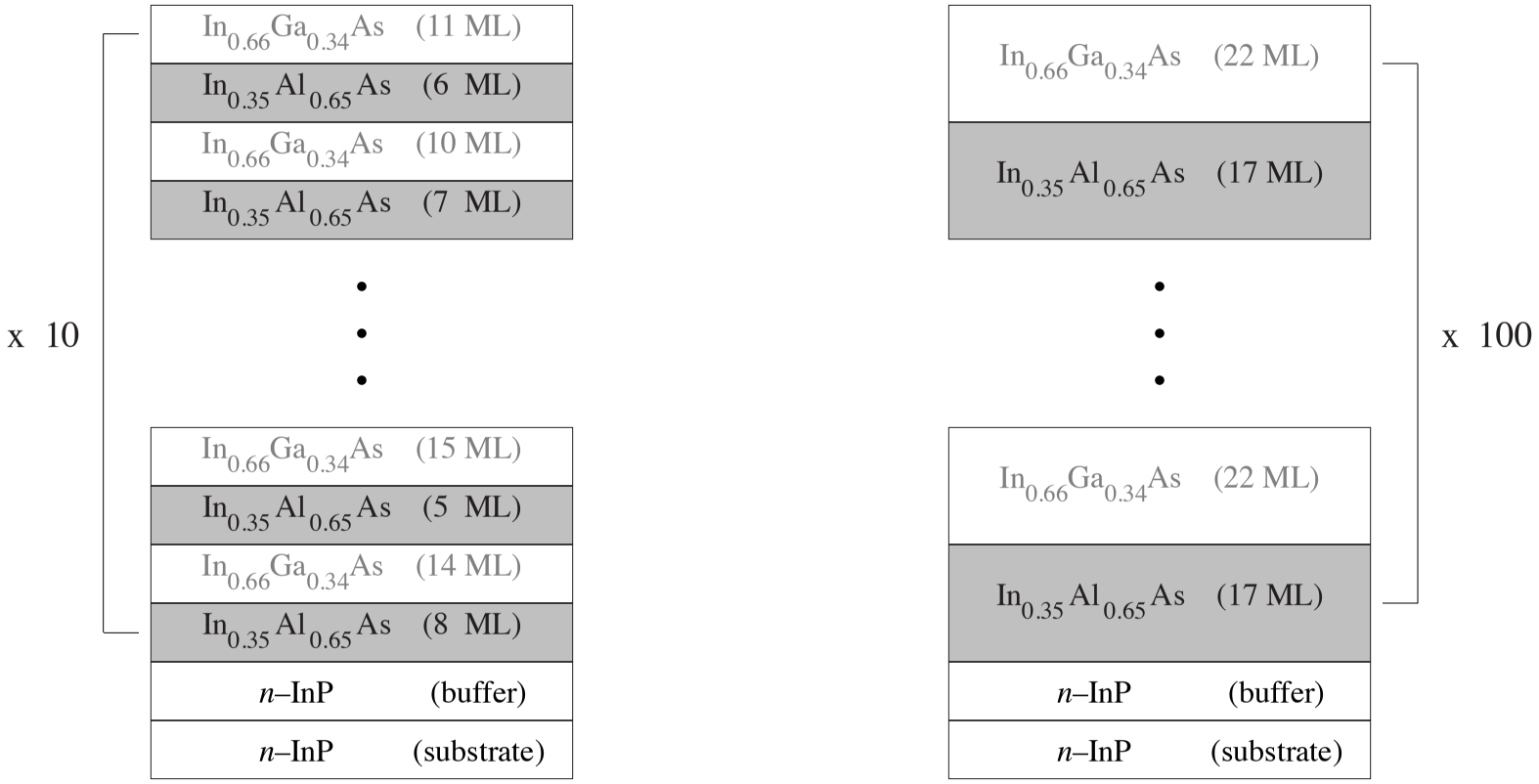


FIGURE 1.6. Schematic layer sequence for a working QCL designed to sense nitric oxide at 5.3 μm (left). A single device stage consists of an aperiodic pattern of twenty–four wells and barriers, only the first and last four of which are detailed here. Ten such stages were grown back–to–back specifically for STM analysis. A simpler, spatially–uniform superlattice (right) with the same strain–balanced stoichiometry and comparatively thick layers, was designed to ease the characterization task. Growth direction is from bottom to top.

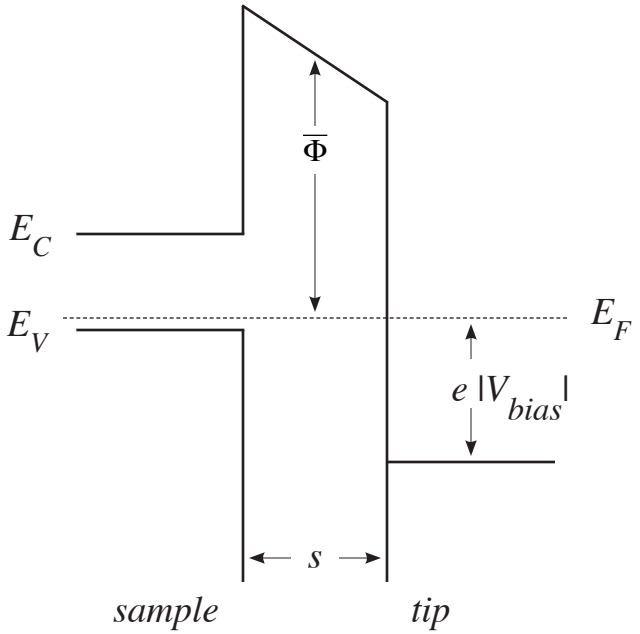


FIGURE 1.7. Schematic electron energy diagram illustrating a trapezoidal tunnel barrier separating an idealized, planar tip-sample junction. Sample is biased, by a negative voltage V, here, to adjust its Fermi level, E_F, with respect to that of the tip. The elementary charge is given by ***e***.

$$I \propto \int_0^{eV} \rho_s(E,eV)\, T(E,eV)\; dE \tag{1.10}$$

corresponding to a convolution of the sample local density of states, ρ_s, with a transmission tunneling probability

$$T(E,eV) \propto \exp(-2\kappa s) \tag{1.11}$$

that depends exponentially on the tip–sample separation, s, and the sample bias voltage, V, through the mean barrier height, $\overline{\Phi}$, according to

$$\kappa = \sqrt{\frac{2m}{\hbar^2}\left(\overline{\Phi}(eV) - E\right)} \tag{1.12}$$

where κ is of the order of 1 Å^{-1}.

The III – V heterostructures of interest in this study adopt the zincblende crystal structure. The non–polar ⟨110⟩ crystal planes, which contain equal numbers of anions and cations, form natural cleavage facets. These ⟨110⟩ facets do not reconstruct following cleavage but instead experience a Jahn–Teller relaxation [28] where anions and cations undergo a coordinated rigid–bond rotation accompanied by a charge transfer that projects filled and empty dangling bonds, respectively, into the vacuum. This charge transfer is conveniently exploited by STM since atom–selective imaging of each III – V material is, under suitable circumstances, achieved by tunneling either out of or into the

surface producing complementary images of the underlying anion or cation sublattices [29].

When an atomically sharp STM tip is brought within a few angstroms of the cleavage–exposed surface, electrons (or holes, depending on the polarity of the bias) from a single tip–atom may tunnel between a single atom at the end of this tip and individual dangling bonds of the constituent surface atoms, thereby providing extraordinary spatial resolution. For example when a III – V semiconductor is negatively biased with respect to the tip, anion–derived dangling bonds that project into the vacuum (such as arsenic or phosphorous) are the dominant source of tunneling electrons. A "topographic" image of these dangling bonds is then constructed line–by–line as the STM–tip is rastered across the surface, following an elevation contour that keeps the tunnel current fixed. In the language of STM, this constant–current contour mirrors a constant electronic–density–of–states of the surface within the energy window available for tunneling [30].

The resulting surface "topography", conventionally represented by monochromatic shades (greyscale) in all of the images described, is influenced by two physically distinct factors. The first arises from actual variations in surface topography – see Fig. 1.8 – where high elevations (protrusions) appear brighter and low elevations (depressions) appear darker. These height differences can be caused by local bond length variations due to alloying or by post–cleavage relaxation within a tensilely – or compressively–strained layer (more in the following chapter). The second factor is due to local variations in the density of states available for tunneling, together with the band–

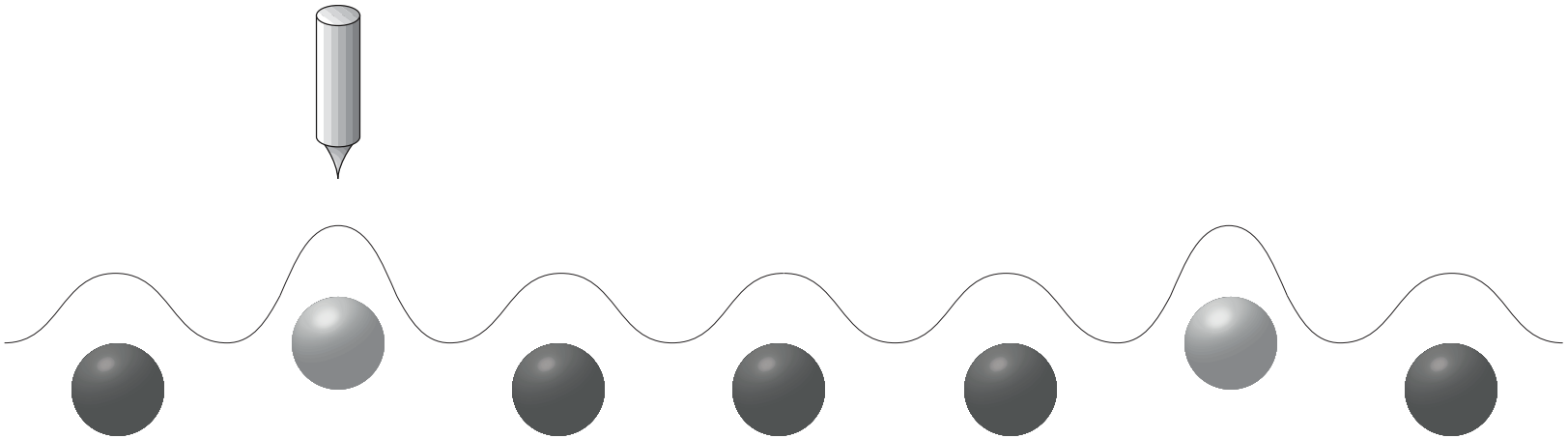


FIGURE 1.8. Schematic illustration of how the tip–sample separation adjusts to maintain a set tunnel current – and a fixed local density of states – in STM over a surface with two, distinct bond lengths. High elevations appear bright, and low elevations dark, in conventional grey–scale images.

edge energy discontinuities (band offsets) introduced during MQW growth, is purely electronic.

The role of band–edge discontinuities is better appreciated by referring to Fig. 1.9 where the band diagrams for a *negatively–biased* InAs / GaSb superlattice are drawn relative to the Fermi energy of a metallic STM tip. Here, tunneling from the sample occurs primarily from carriers near the top of the respective valence bands. For a fixed tip–sample voltage, electrons originating from InAs (electron well) experience a lower vacuum barrier than those originating from GaSb (electron barrier) due to the valence band offset between the two materials. Because InAs and GaSb are very nearly lattice matched, there is no (perceptible) bond–length discontinuity at the cleavage surface between the two materials and the respective change in tunneling current depends only on the vacuum barrier. The GaSb layers will therefore appear higher (brighter) than the InAs layers in filled–state [31] (i.e., negative–sample–bias) images as the tip–sample separation adjusts to maintain constant tunneling current.

A particularly instructive example is offered by the filled–state STM image of an InAs / GaSb superlattice shown in Fig. 1.10 [32]. This image serves, first, as the epitome of state–of–the–art STM technique achieved in our laboratory prior to investigation of InP–based QC materials. This image also highlights the contributions to the STM–derived contrast from distinct bond length variations together with the effect of energy–band discontinuities (from the constituent materials). Finally, this STM image illustrates how individual isovalent impurities can be identified, categorized, and counted monolayer–by–monolayer in the growth direction to completely characterize a given

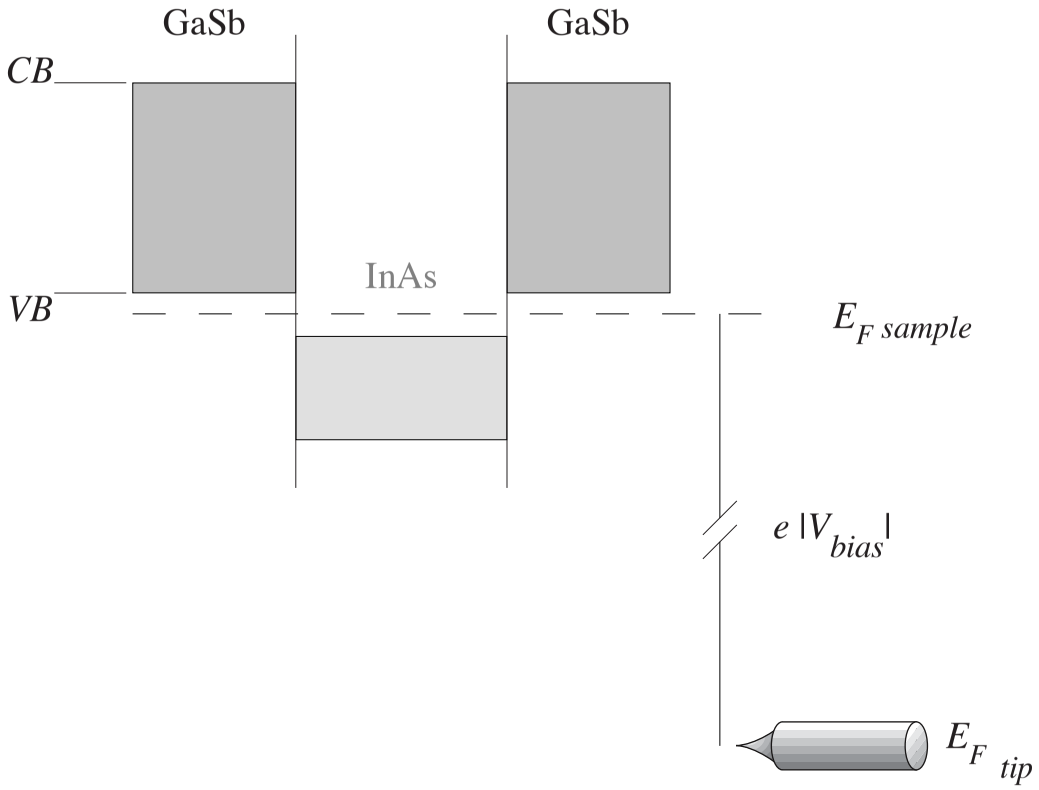


FIGURE 1.9. Energy band diagram for a InAs / GaSb type–II superlattice illustrating electronic contribution to observed STM layer contrast from band discontinuities in MQWs. GaSb layers appear brighter than InAs layers in filled–state STM images due to the higher probability for electrons to tunnel through the vacuum barrier. The STM tip is illustrated as if tunneling occurs at the (001) growth surface to emphasize the relative energy alignments; in reality, the tip is rastered left to right across this diagram in a perpendicular {110} surface.

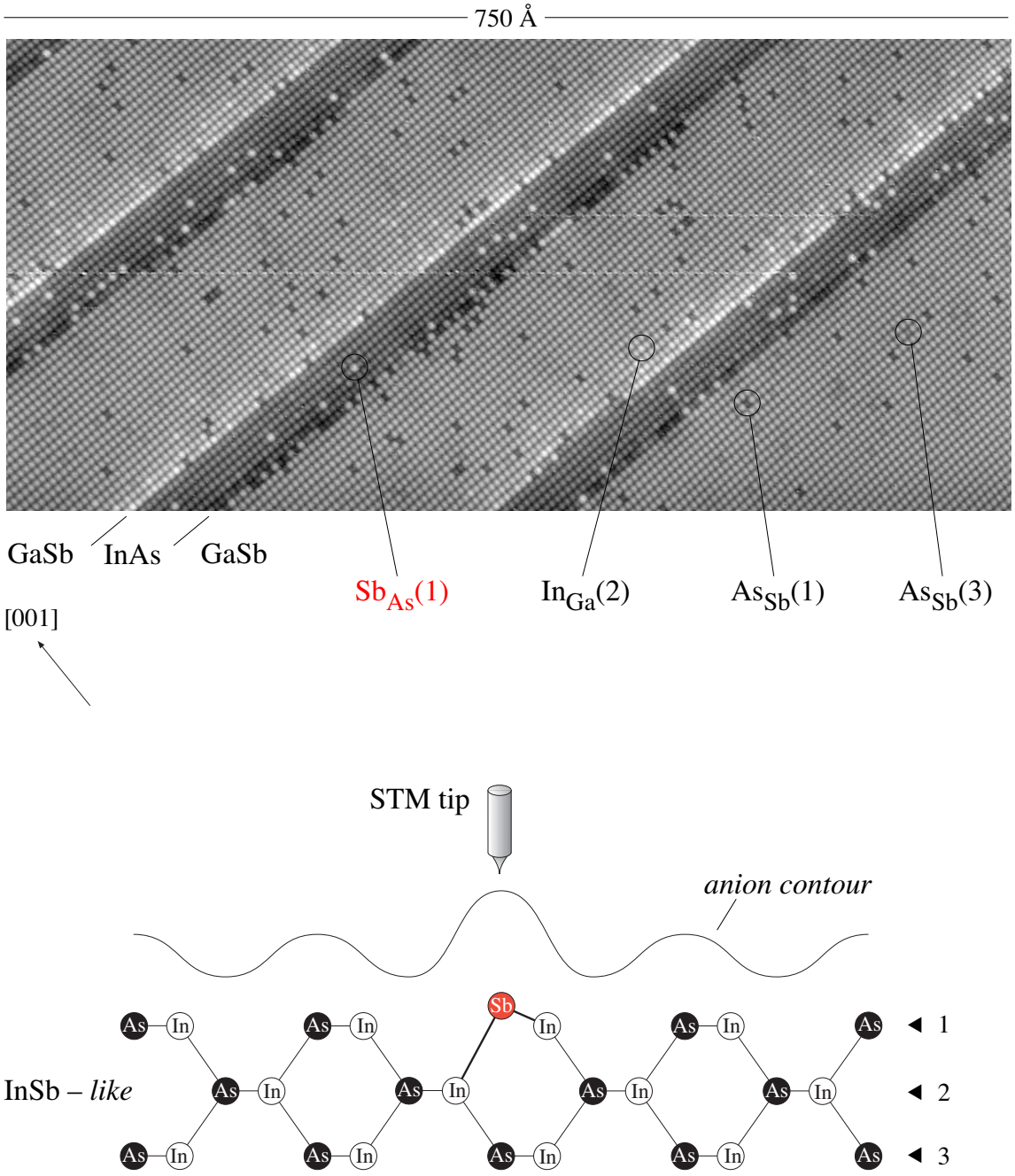


FIGURE 1.10. Atomic–resolution STM image (top) of the anion sublattice (Sb, As) for a type–II InAs / GaSb superlattice. Isovalent impurity substitution in the surface (1) and sub–surface (2,3) planes are indicated in the image. Electronic contribution to the observed layer contrast is manifested as GaSb layers appearing brighter than InAs layers. Growth direction is from bottom–right to top–left. Schematic anion contour (bottom) across an isolated InAs layer highlighting an antimony–for–arsenic isovalent substitution and associated change in bond length. Growth direction is left to right. Reprinted with permission from [32].

growth [33].

This real–space approach comes natural for a non–common–atom (InAs / GaSb) system where individual atom identification is readily available, however in a more homogeneous common–atom (InAlAs / InGaAs) structure, such as those found in typical QC lasers, the absence of isolated features – see Fig. 1.11 – prevents atom identification thereby removing the utility of direct atomic counting. It is for this particular reason as well as the fact that we seek the characterization of continuous interfaces with cross–sectional STM that we must rely on spatial insights derived from reciprocal–space discernments.

Leaving behind InAs / GaSb structures, the studies presented here will concentrate only on the examination of InAlAs / InGaAs QC materials. A typical filled–state STM image from the active region of a working QCL grown by MOCVD is shown in Fig. 1.11. Here one notes the layer contrast contributed by energy–band discontinuities and post–cleavage strain relaxation (more to follow in the next chapter) is visibly growth direction dependent. The influence of the complex layer geometry inherent to QCL designs described earlier on the observed STM contrast –unmistakable in this STM image – likewise motivates our adoption of a simplified *periodic* design (with comparatively thick and uniform layers throughout) for the purpose of analyzing interface roughness.

Epitaxial growth of the structures examined here was carried out at Corning Incorporated with a Turbodisc D180 vertical MOCVD reactor manufactured by Veeco and operated at 70 Torr. Epitaxial layers were deposited atop an (001)–oriented, n–type

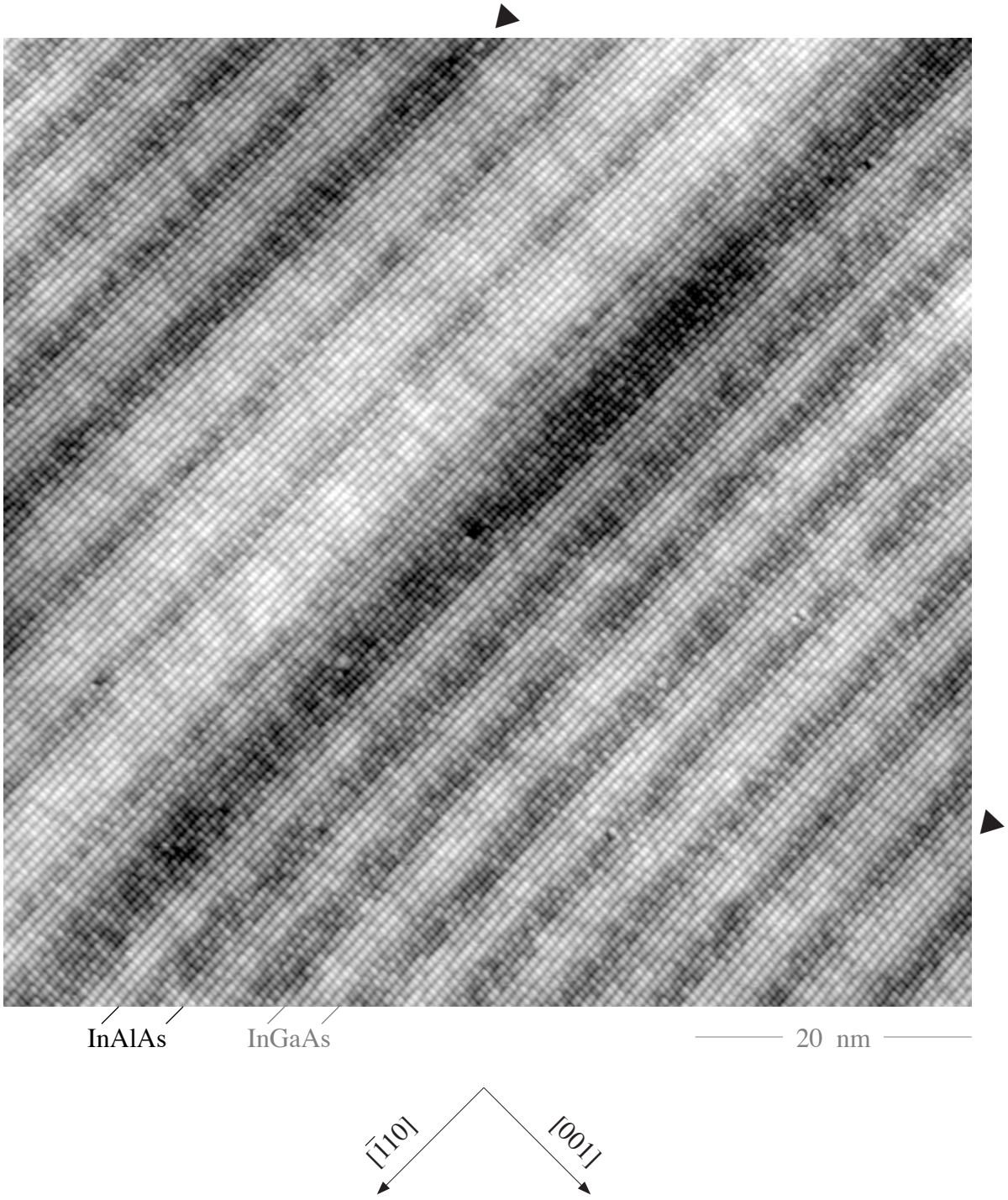


FIGURE 1.11. Atomic–resolution STM image of the arsenic sublattice throughout one of the ten, QCL stages (grown by MOCVD) illustrated in Fig. 1.6. The geometric complexity of the structure is evident. Top–left and bottom–right carets demarcate first and last layer of such QCL stage, respectively. QC device courtesy of Dr. C. Caneau, Corning Incorporated. Growth direction is from top–left to bottom–right.

InP substrate which was rotated during growth at a temperature of 600 ºC with growth rates of 5 Å / sec and an approximate anion–to–cation ratio of 200. The QC device was designed to monitor and sense nitric oxide at a wavelength of 5.3 μm [34].

Dissertation Overview

Remaining chapters of this manuscript will cover in the following order: an overall description of technical challenges – and corresponding resolutions – encountered during sample–cleavage exploration and the monumental UHV system reconstruction necessary to maintain a pristine environment suitable for freshly cleaved aluminum–bearing materials; an overview of well–known non–idealities present in piezoelectric scanning probes and our workarounds to mitigate these inherent deficiencies; the development of a novel method to accurately measure superlattice periodicity with cross–sectional STM; the detailed steps necessary to digitally extract heterojunctions from STM images; a systematic examination of the effects that Gaussian correlations exert on roughness measurements; finally the extraction of correlation lengths from experimentally obtained interface profiles, and a critical discussion of our findings.

CHAPTER II

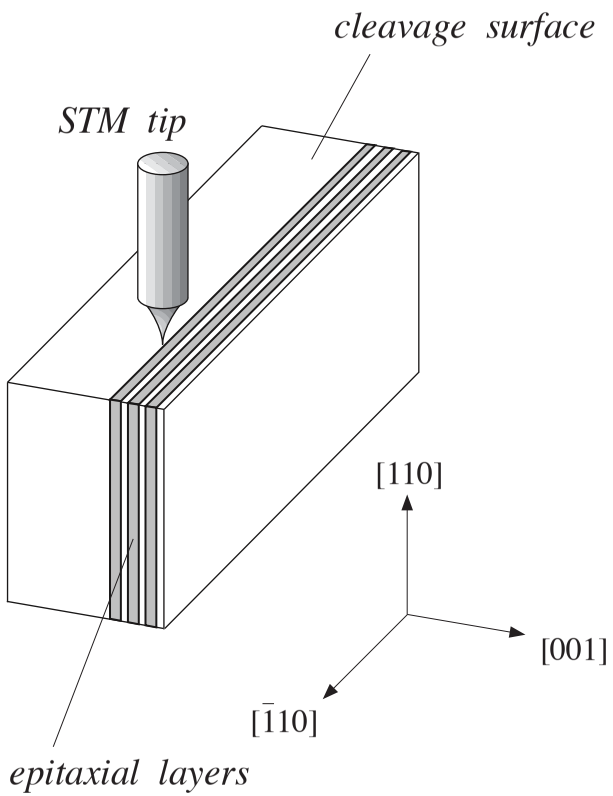
EXPERIMENTAL DETAILS

Introduction

The STM employed in these studies is a commercial unit manufactured by Omicron Vacuumphysik [35]. Details regarding technical specifications of the microscope may be found in earlier dissertations [36, 37].

The STM resides in a dedicated UHV chamber – see Fig. 2.1, right – that provides a nominally pristine environment to ensure freshly–cleaved semiconductor surfaces remain contamination free (Fig. 2.1, left). As depicted, this instrument is the product of a prolonged and meticulous re–engineering of an inherited vacuum system that, although superb by ordinary standards, just was not up to the demands posed by such a highly–reactive material as InAlAs. In short, QCL samples cleaved in the inherited vacuum environment became too contaminated for imaging with STM in a matter of hours; following the arduous upgrade from UHV to extreme–high vacuum (XHV) conditions described below, these cleaves would remain useable for days.

Further technical challenges introduced by this “new” material system posed similarly daunting barriers to progress. Indeed, these technical challenges were so numerous, and so formidable, that prospects for successfully achieving the project's already ambitious scientific objectives were cast into doubt more than once. Practically speaking, every step, from sample preparation to data acquisition, presented new and



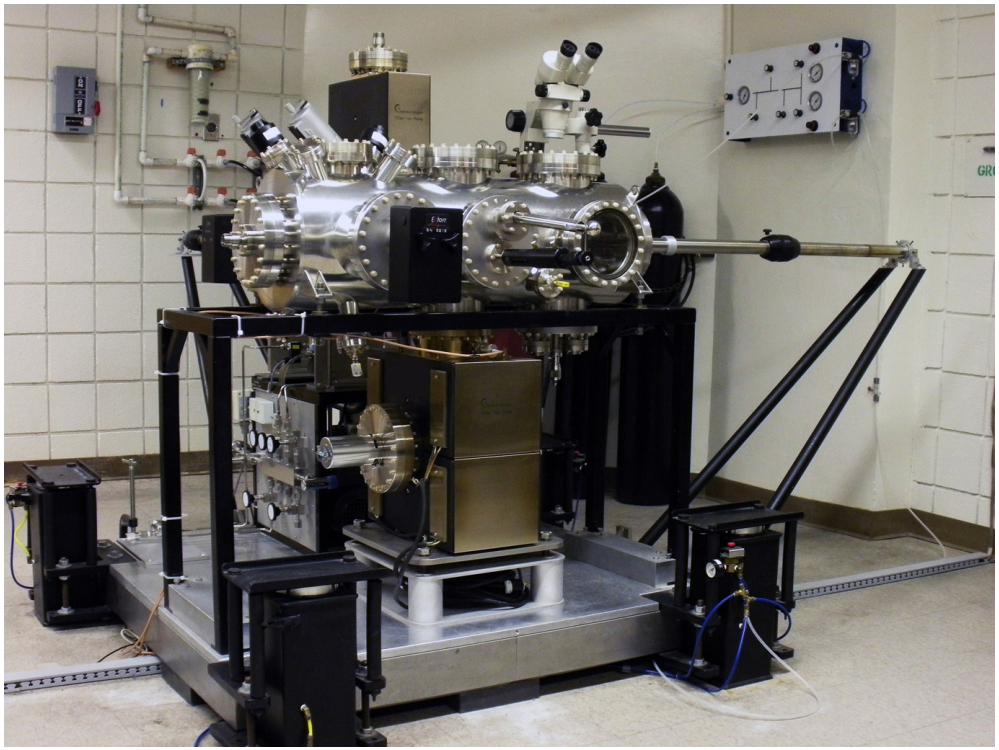


FIGURE 2.1. Illustration of the crystal cross sections accessible via in–situ cleavage (left), alongside the dedicated UHVchamber (right) where the scanning tunneling microscope resides.

unique demands that, one by one, had to be methodically and creatively addressed.

Technical Challenges

The Texas A&M STM laboratory previously enjoyed a long and successful history in applying cross–sectional STM to GaSb–based type II QWs [20, 21, 38 – 41] where the large–area, mirror–like, cleavage facets needed for atomic–resolution imaging are reproducibly obtained. When the same, previously successful, cleavage protocol was applied to InP–based samples, however, it soon became clear this substrate possesses altogether different fracture characteristics. That difference is best appreciated by referring to Fig. 2.2, which presents the Nomarski micrographs from orthogonal cross–sections of InP – (110) and (1–10), the two natural cleavage planes for any zincblende semiconductor – using the inherited technique.

The superlattices ultimately studied here were all grown on two–inch InP wafers. These wafers are then diced into small 5 mm x 5 mm squares with an automated laser scriber in our laboratory to facilitate their subsequent vacuum cleavage and mounting in the STM; the edges of these dies are chosen parallel to major and minor wafer flats as indicated in Fig. 2.3 The laser scriber is likewise used to score a small notch near the center of each die that is again aligned with one or the other of these flat directions; this notch serves to initiate, and preferentially direct, subsequent fracture along either a (110) or (1–10) plane.

Since our initial attempts at cleaving InP substrates (Fig. 2.2) proved so disappointing, the entire protocol was revisited from beginning to end, with each

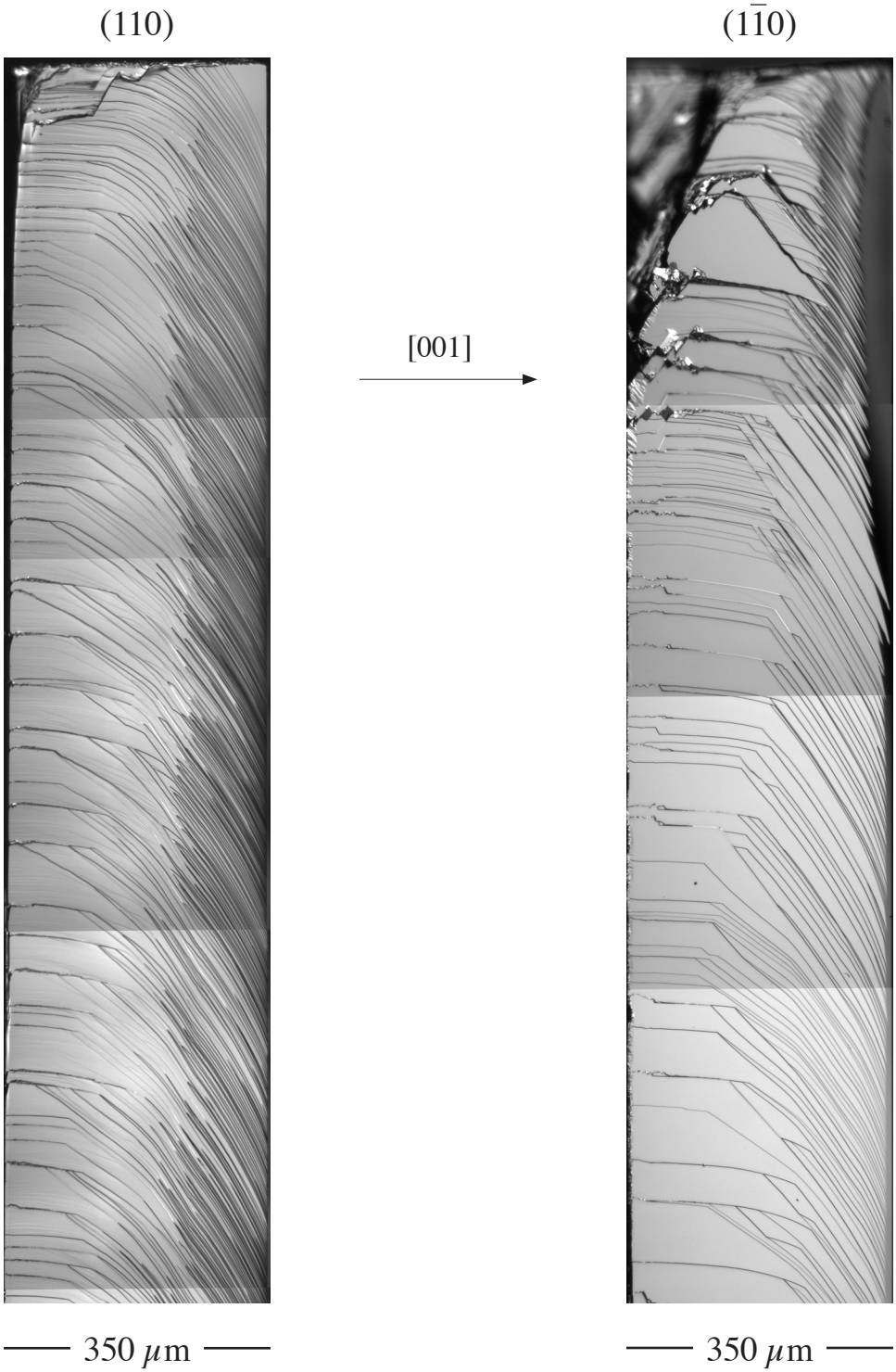


FIGURE 2.2. Nomarski micrographs exhibiting typical cleavage results achieved for InP substrates in either (110) (left) or (1–10) (right) cross sections when employing inherited cleaving protocols developed for GaSb–based semiconductors. Growth direction is from left to right.

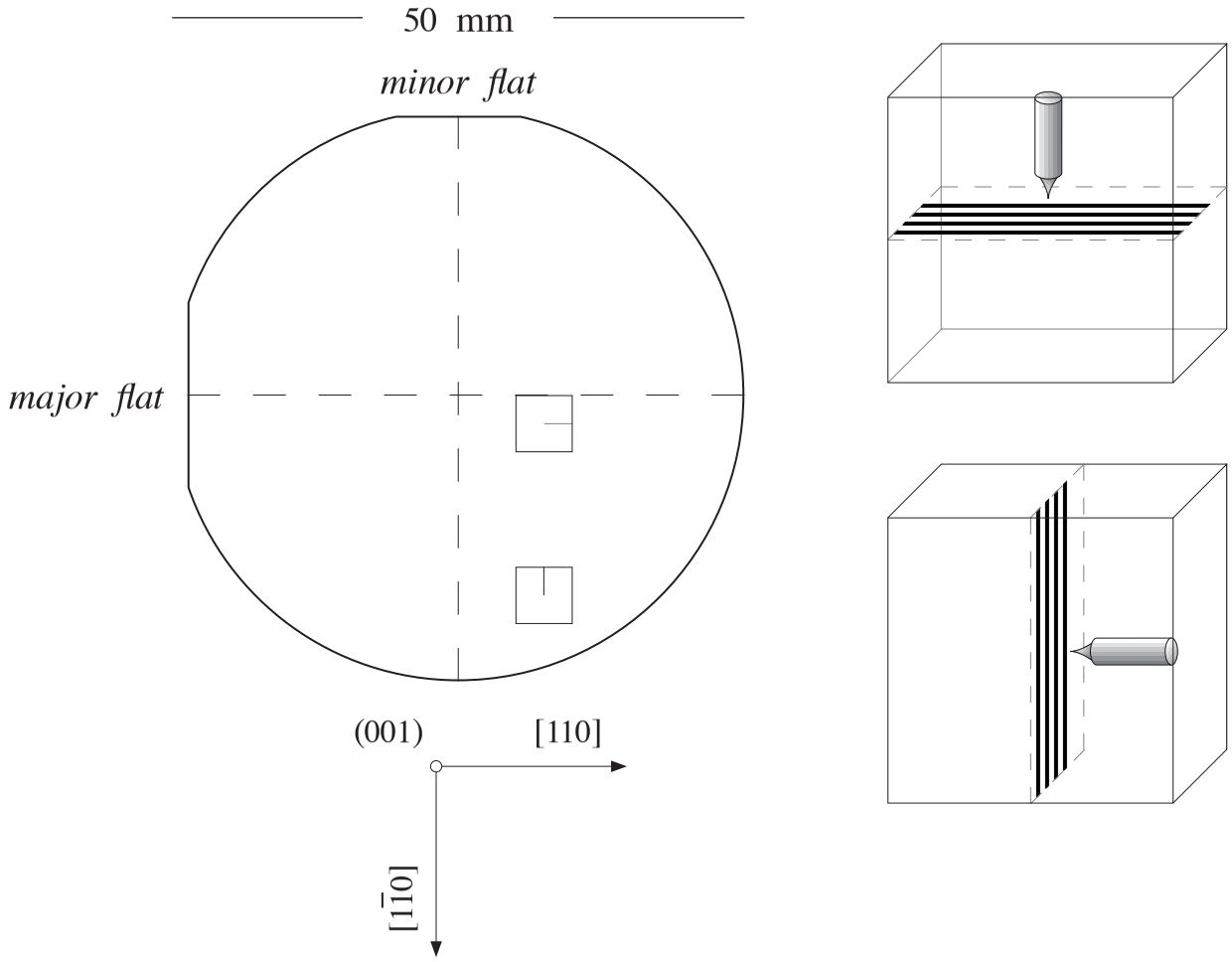


FIGURE 2.3. Schematic illustration of a commercial "EJ" oriented semiconductor wafer with crystal planes indicated by conventional major– and minor– flats. Fracture process can be directed parallel to either a major– or minor– flat via a small notch aligned with the corresponding crystal axes. STM tip is shown here as if it is rotated to probe the exposed epilayers in orthogonal cross sections, in reality the tip orientation remains fixed while the sample is rotated.

component of the existing process (direction, length, and depth of the fracture–initiating notch; sample mounting; impact location and impact speed) systematically varied, keeping the remainder fixed as though each variable was independent of all others. This exploration was exhaustive, requiring over a year of iterative, trial and error before optimum cleavage conditions were convincingly established. The following paragraphs recount key points of this all–too–long process, but our exposition serves as a clear narrative rather than a chronologically accurate account.

Figure 2.4 summarizes the major steps comprising our cleavage protocol, whose consequential details are further explored in Figs. 2.5 and 2.6. We begin with Fig. 2.4 (a), which reveals a surprising but nevertheless important finding: the sense in which the laser beam is rastered over a die to produce the desired cleavage notch (Fig. 2.3) matters profoundly. Notches scored in either crystal direction (Fig. 2.3) with center–to–edge rasters yield notably better cleaves than the reverse, presumably due to some (uncontrolled) latency in the laser–beam stepper motors at, or near, the conclusion of each raster line.

Once a die has been notched in the desired direction, it is mounted in a specially–designed sample holder where it is securely fastened between a stainless–steal anvil and stainless–steel clamp, as shown in Fig. 2.4 (b). A beryllium–copper spring, inserted to one side of the clamp, provides a source of stress relief that prevents sample destruction due to differential thermal expansion during vacuum bake out. The STM die and BeCu spring were historically centered between the two fastening screws of the sample holder as suggested in Fig. 2.4 (b); it turns out (again surprisingly) that asymmetrical vertical

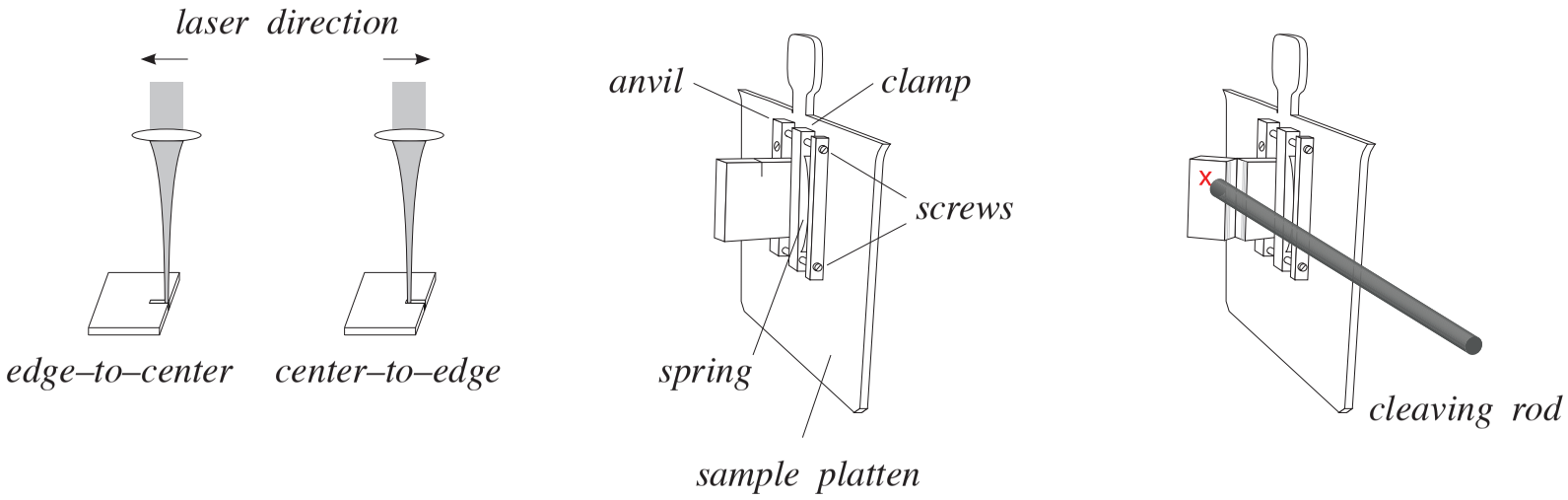


FIGURE 2.4. Overview of major steps systematically revisited to modify inherited substrate cleavage protocols for optimum results with InP. Scribing direction (a), sample mounting (b), and impact location (c), each play important roles.

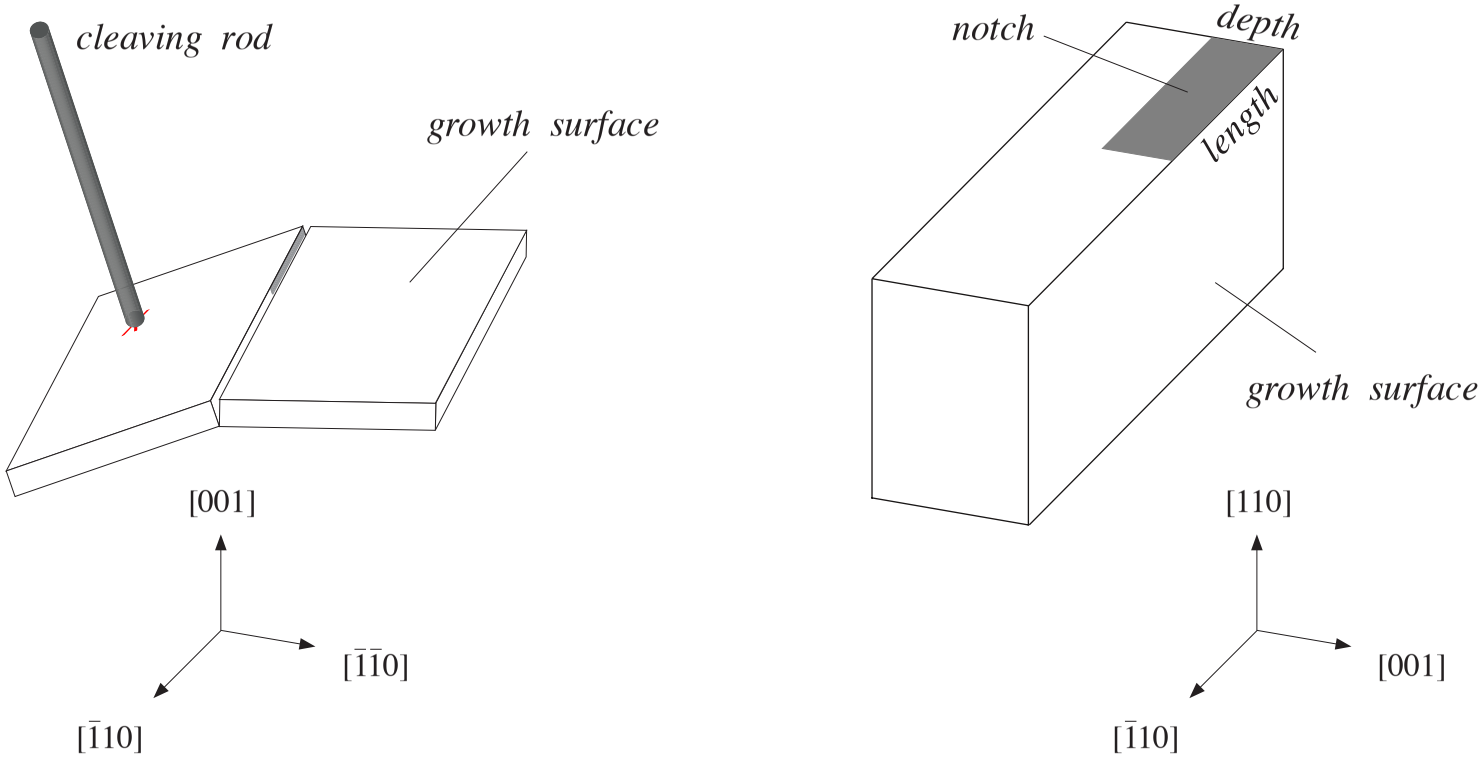


FIGURE 2.5. Rotated and exploded view (right) of exposed substrate notch morphology (in cross section) following sample–die cleavage (left). Notch length and notch depth are both important factors, with the former relatively easy, and the latter comparatively difficult, to control.

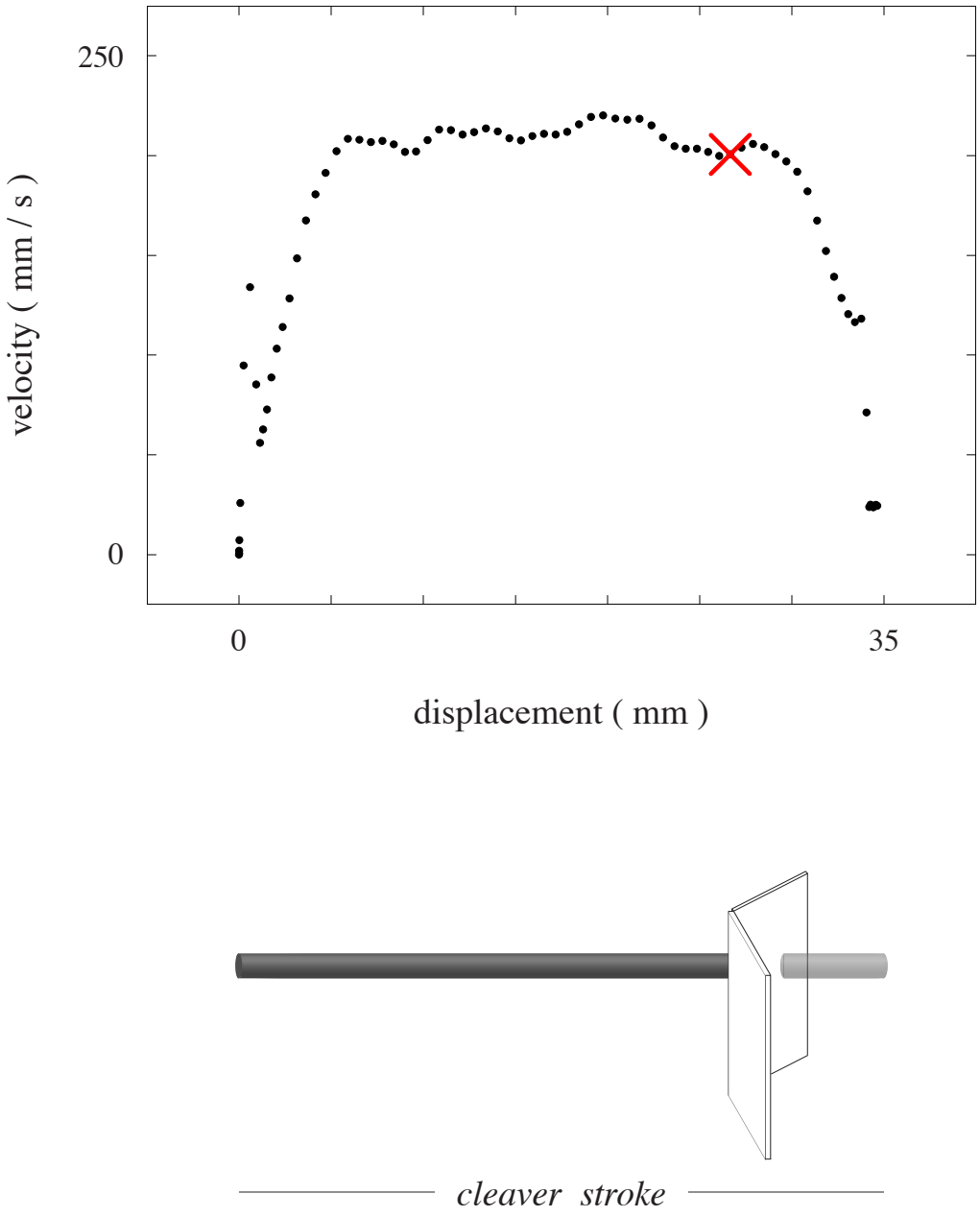


FIGURE 2.6. Pneumatically–actuated cleavage rod velocity versus displacement (top) and accompanying real–space stroke schematic (bottom). The velocity at impact (X) may be read directly off this graph once the fixed, retracted–rod–to–sample distance is known.

positioning of the STM die and BeCu spring – with the die placed as low as the bottom screw will permit and the BeCu spring at or very near the top screw – provides consistently better cleavage results.

Once firmly mounted to its sample holder, a notched die is ready to sustain an impulsive impact that exposes its buried epitaxial layers, and laser–induced notch morphology, in cross section. This impact is most effectively positioned near the top–left corner of the discarded–half die, as indicated in Fig. 2.4 (c).

The situation in Fig. 2.4 (c) is pursued more carefully in Fig. 2.5, where the role of cleavage–notch morphology – the length and depth of the laser score, in particular – is emphasized. Control over the length of a center–to–edge raster is easily effected via the laser–beam stepper motors once this stage has been appropriately aligned with the laser beam; control over the depth of a notch is a more challenging matter, however.

The damage sustained by any semiconductor during high–intensity laser illumination depends on a complicated interplay of beam parameters (intensity, wavelength, beam profile, depth of focus, etc.) and material properties (absorption coefficient, thermal conductivity, etc.). However, for a specific beam profile and wavelength (1.06 micron), as well as a given substrate material (*n*– or *p*–type InP) with known thermal and optical properties (absent significant perturbation by additional epilayers), only the total energy delivered to each unit area of the surface matters. This energy is the product of two factors: beam intensity per raster point (watts per unit area) and dwell time per raster point (time per raster step).

Reproducibility suggests beam power and beam profile remain fixed, within the accuracy limits attainable with standard calibration procedures, while the step–time increment of the laser–beam stepper motor is manipulated to control the dwell time at each point. It was found, in this way, that the best cleavage results for InP are typically achieved with a notch length of ~1.8 mm, and notch depth of ~100–120 micron, or roughly one–third the original wafer thickness.

Figure 2.6 emphasizes another aspect of Fig. 2.5 that plays a similarly important role: the velocity of the cleaving rod as it strikes the growth surface. This rod is pneumatically actuated [42] to provide a reproducible stroke that is unattainable otherwise. The linear acceleration of the rod is monitored (as a function of time) with a commercially available USB accelerometer [43] fastened to the atmospheric side of the actuator. The measured acceleration is converted to a linear velocity by discrete integration of the data, and a subsequent integration of this velocity yields the rod's linear displacement. A plot of velocity versus displacement so obtained (Fig. 2.6) then permits the velocity at impact position (X) to be determined from the known travel distance to the sample. Collective experience indicates actuator speeds in the range of 200 – 300 mm / s produce the best cleavage results with InP.

Figure 2.7 summarizes the extraordinary progress thus painstakingly achieved with a side–by–side comparison of typical fracture morphologies from inherited and improved cleavage protocols.

The STM instrumentation resides in a dedicated UHV chamber that provides a pristine environment to ensure a contamination free *in–situ* habitat for freshly–cleaved

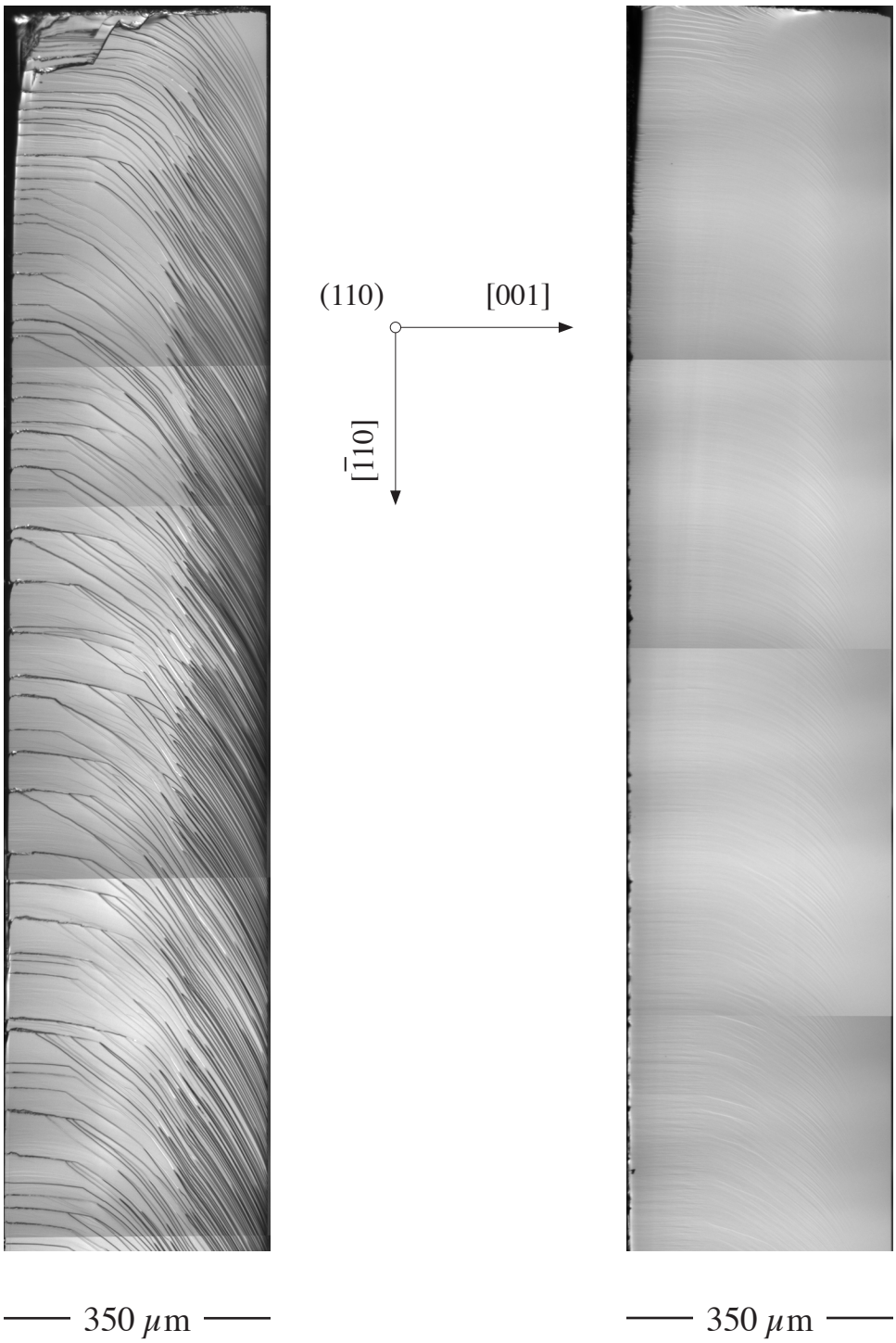


FIGURE 2.7. Nomarski micrographs contrasting (110) InP facets obtained with inherited (left) and improved (right) substrate cleavage protocols. Similar results are obtained for (1–10) InP facets. Growth direction is from left to right.

semiconductor surfaces. Early experiments made it clear the InAlAs barriers in a typical QCL structure reacted too rapidly with the already–low, residual–gas levels in the existing UHV chamber to successfully pursue the scientific studies contemplated here.

A typical STM image obtained from an InAlAs / InGaAs superlattice with the chamber in its inherited state is shown in Fig. 2.8; this image was acquired approximately twenty–four hours post cleavage. The preferential agglomeration of (bright) contaminants over the (dark) InAlAs layers is unmistakable, so that contamination evidently takes place selectively, rather than uniformly, over the exposed surface. A moments reflection suggests this should come as no surprise: the typical residual gas species in a stainless–steel UHV system are, hydrogen aside, oxygen–bearing Lewis acids (electron donors) that will very quickly react with aluminum, one of the strongest known Lewis bases (electron acceptor).

In order to reduce all aluminum–reactive species such as water, carbon monoxide, carbon dioxide, and oxygen to suitable levels, the entire vacuum apparatus was subjected to a painstaking transformation. A front view of the main chamber (housing the STM) in its inherited state is shown in Fig. 2.9. Nearly all of the components originally attached to, or installed within, this chamber were ultimately removed or upgraded. The stainless–steel shell and STM (together with its associated cleavage stage) are all that remain of this original setup.

Reconditioning began with the elimination of all unnecessary instrumentation to immediately reduce the outgassing burden on the system's vacuum pumps. For example, the main chamber was originally equipped with an Auger spectrometer and triple–axis

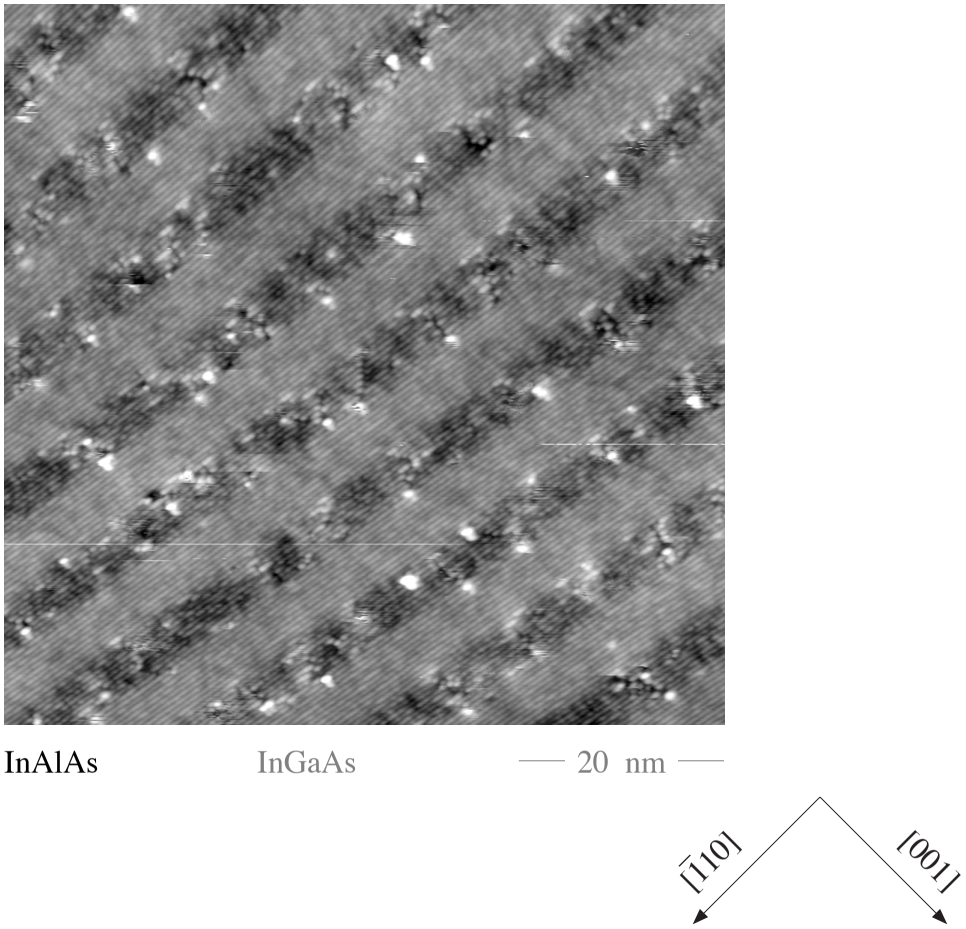


FIGURE 2.8. Typical STM image from a freshly–cleaved InAlAs / InGaAs SL following 24–hour exposure to the ambient environment of the original (left) UHV system. Contamination – observed primarily within the InAlAs (dark) barriers – is due to oxygen–containing (CO, CO_2, O_2, H_2O) residual gas species.

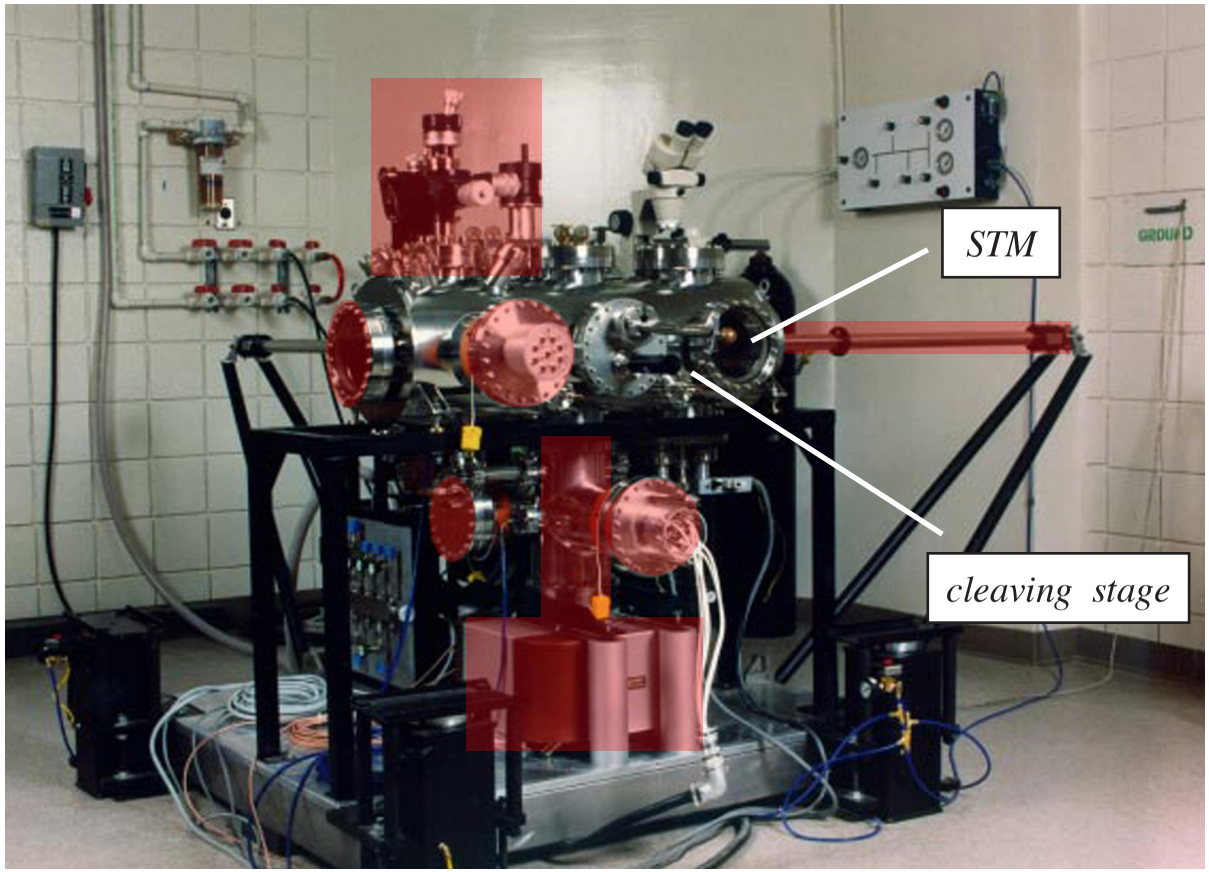


FIGURE 2.9. Front view of the original UHV system (ca. 2006) prior to major reconditioning. Components highlighted in red were permanently removed. The only components that remain from the original setup are the main chamber, cleaving stage, and STM.

(XYZ) sample manipulator that were no longer relevant and thus removed. At the same time, a well–worn, magnetically–coupled, linear–rotary feedthrough routinely used to transfer freshly cleaved samples from the cleaving stage to the STM (Fig. 2.9) was replaced with a newer design that was smoother acting, more easily pumped, and exhibited reduced outgassing of the reactive, oxygen–bearing species of concern.

These modifications notwithstanding, the outdated pumps used to generate and sustain the UHV environment of the inherited system proved incapable of achieving the XHV levels ultimately sought, so they, too, had to be replaced with an entirely new generation of more–powerful models. The upgrade, here, necessitated a carefully–thought–out combination of sputter–ion pumps, non–evaporable getters (NEG), and titanium sublimation pumps (TSP) with complementary getter characteristics. Fig. 2.10 shows the main chamber in this upgraded state, with the new components, including a state–of–the–art residual gas analyzer (RGA), individually identified.

Another pair of chambers, connected to the main chamber via a system of manually–actuated gate valves, serves as an entry point for the insertion of samples and tips from atmospheric pressure into the XHV environment of the STM, as well as their subsequent retrieval. This entry system, perpendicular to the main chamber and therefore obscured from the viewpoint of Fig. 210, consists of two stages: a stainless–steel load lock, together with an intermediate, stainless–steel transfer chamber isolating the main chamber from the load lock. The entry system is depicted in its inherited state in Fig. 2.11.

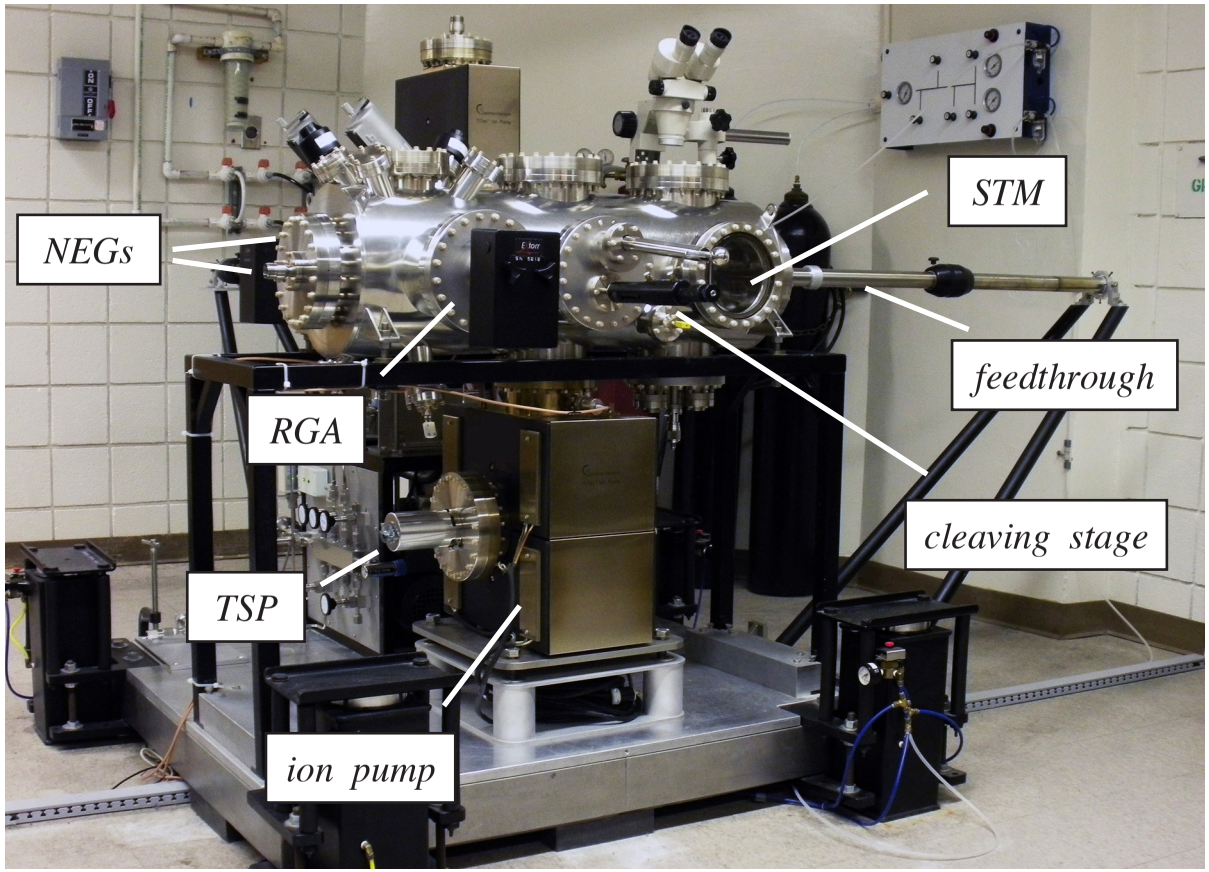


FIGURE 2.10. Front view of the upgraded UHV system (ca. 2008) post major reconditioning. Newer, more powerful vacuum pumps and a new smoothly actuated linear feedthrough were installed in the main chamber to provide a suitable habitat for freshly cleaved Al–rich surfaces. Vacuum improvements were monitored with a new electron–multiplier enhanced RGA.

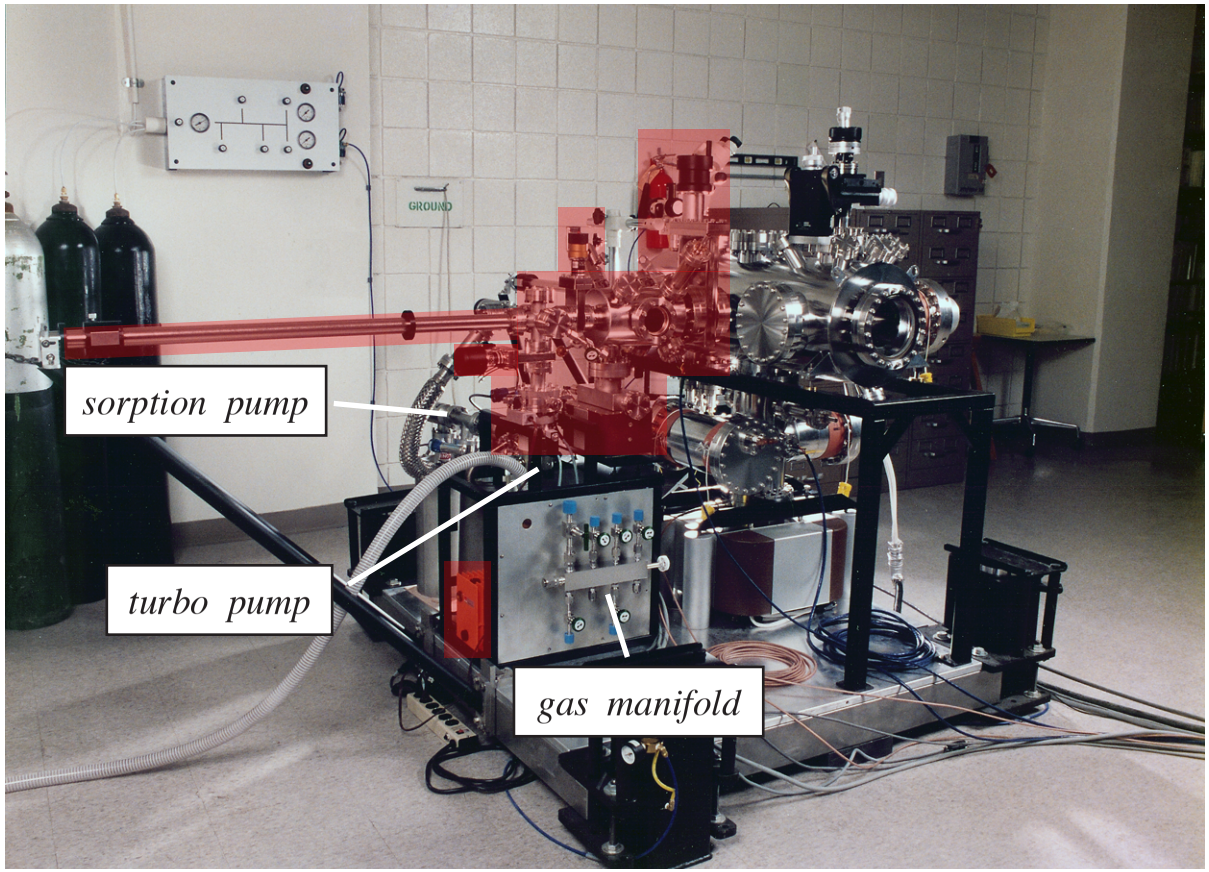


FIGURE 2.11. Side view of the original UHV system (ca. 2006) prior to major reconditioning. The sample introductory system – highlighted in red – was replaced with an entirely new load lock and transfer chamber design. The only original components still in use are the load luck pumps (turbo molecular, sorption) and gas manifold.

The need to recondition the entry system was dictated by circumstances altogether different from those pertinent to the main chamber: the entry chamber suffered devastating, irreversible contamination following the unintentional introduction of a cadmium–containing component some years earlier. The affected setup had to be entirely dismantled, and a new pair of chambers designed and constructed.

The new entry design incorporated several crucial improvements, foremost amongst them the installation of a strategically–located residual gas analyzer (RGA) within the load–lock itself to immediately detect, and isolate, any future contamination. In addition, the minimum diameter for all stainless–steel fittings connected to the main chamber was increased from 1.5" to 4.0", considerably improving gas throughput to the external, turbo–molecular pump during system bakeout, and thereby improving ultimate vacuum. Incorporation of a new–generation ion–pump / NEG combination in the transfer chamber likewise succeeded in damping disruptive, transient gas loads otherwise experienced by the main chamber during sample introduction and retrieval. Finally, the favorable vacuum, and mechanical, characteristics described in connection with the magnetically–coupled linear / rotary feedthrough upgrade chosen for the main chamber likewise dictated its adoption, here, for the entry system. Fig. 2.12 shows the entry system in its present, upgraded, configuration, with each of the new components individually identified.

As alluded to previously, mechanical motion in vacuum introduces outgassing transients over and above the static outgassing burden that (in conjunction with pumping speed) determines ultimate pressure. Since these transients exert a proportionately

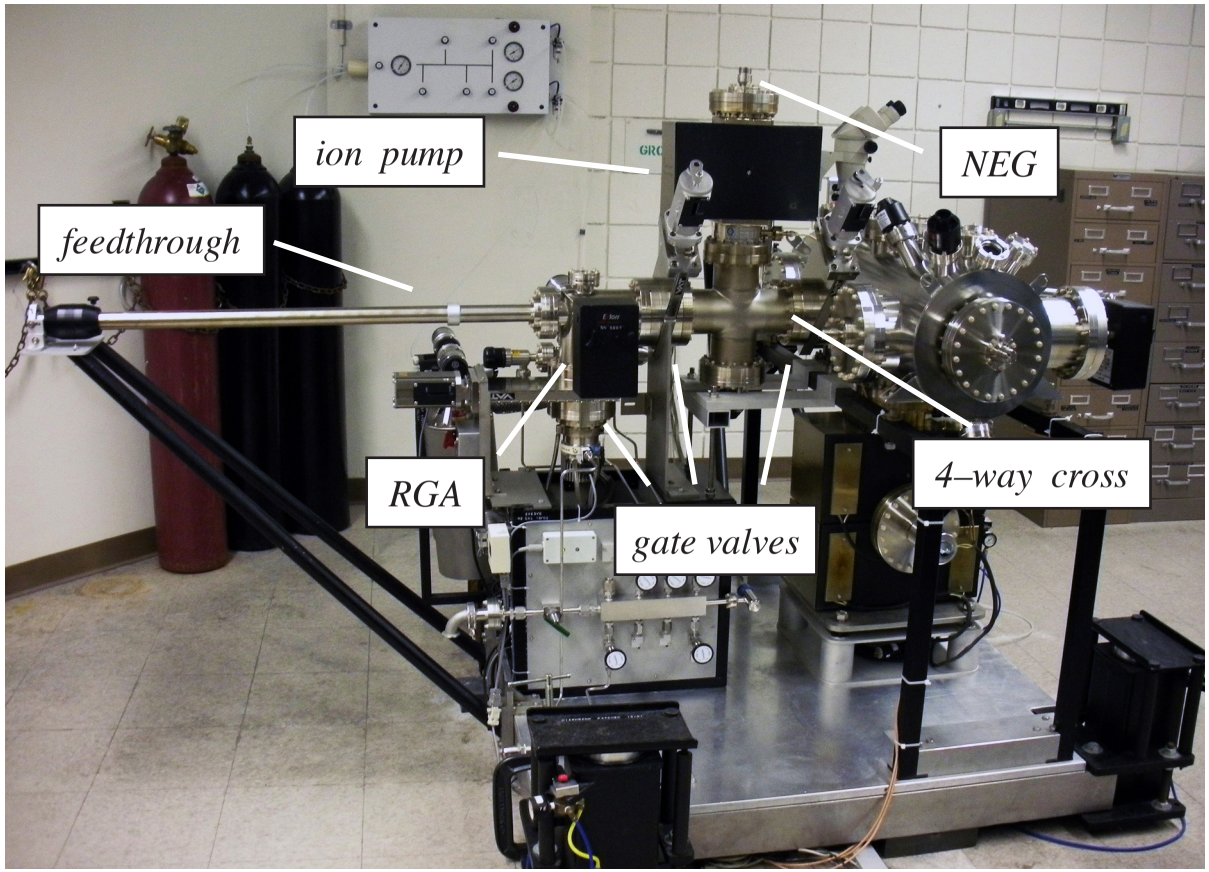


FIGURE 2.12. Side view of the upgraded UHV system (ca. 2008) post major reconditioning. Vacuum pumps, gate valves, and the linear motion feedthrough were all replaced with newer, high–performance components; in addition, a dedicated, electron–multiplier–enhanced RGA was added to the sample load lock. Stainless–steel tubulation increased from 2.75" to 4" OD minimum throughout, resulting in lower, transient gas loads to the main chamber during sample insertion or retrieval, as well as a lower ultimate pressure following system bakeouts. A simple, 4–way cross serves as a transfer chamber.

detrimental effect on freshly–cleaved InAlAs surfaces, the desire to minimize transport and / or manipulation of such samples following cleavage is self–explanatory. An important, but previously overlooked, observation directly bears on this point: an (unavoidable) outgassing transient accompanies every (avoidable) actuation of the cleaving–stage carousel routinely used to re–orient the sample holder for subsequent mounting onto the STM. Specifically, with the inherited sample–mounting setup (Fig. 2.13), the surface normal following cleavage is parallel to the (fixed) STM tip; a freshly–cleaved sample must then be rotated so this surface normal is antiparallel to the tip and the two directly face one another (Fig. 2.13). That rotation can be avoided, thus eliminating any post–cleavage carousel actuation with its attendant pressure transient, by cleverly mounting samples to the backside (rather than front) of the sample platten, as shown in Fig. 2.13.

The log–log plot of RGA partial pressures versus time in Fig. 2.14 (left) demonstrates the pristine, sub–pico Torr environment attained with this state–of–the–art system for every residual–gas species save hydrogen[1]. These measurements may be recast into more tangible and easily visualized terms by plotting the corresponding monolayer formation time, assuming unit sticking probability, for each reactive component; this is shown in Fig. 2.14 (right). Under ideal, steady–state circumstances, then, the expected surface lifetime is CO–limited and of order 200 days. Since sticking coefficients are rarely unity, pressure transients during sample transport cannot be

[1] The data in Fig. 2.14 were assembled over a period of two weeks, and the measurements were conducted under steady–state conditions, without permitting actuation of any in–vacuo mechanical components.

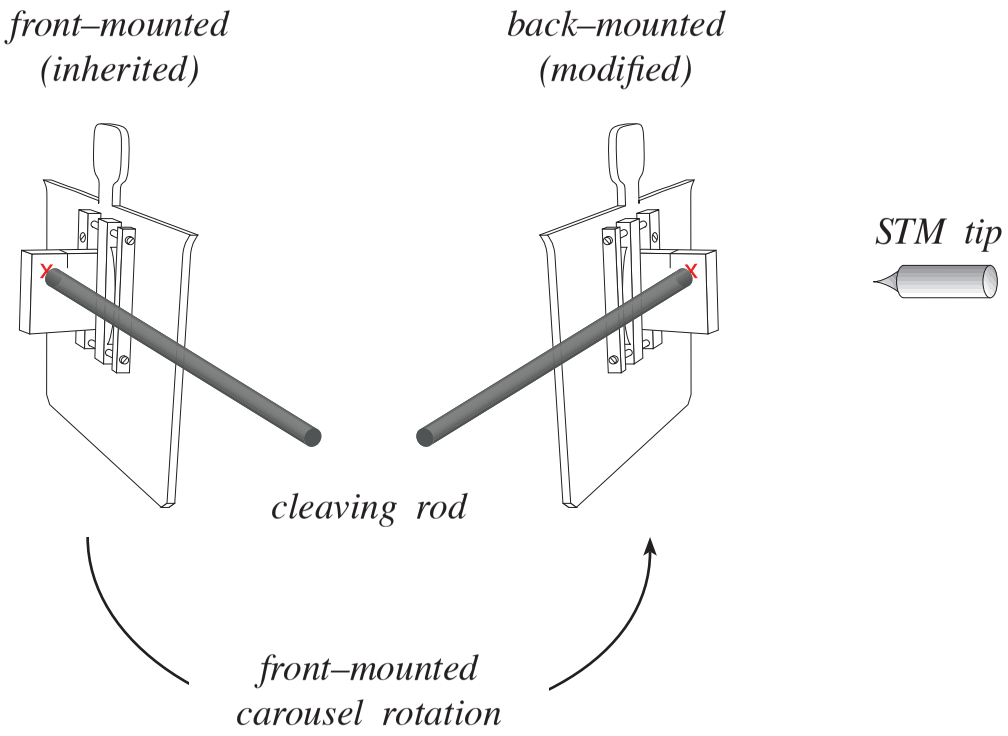


FIGURE 2.13. Traditional fastening of scored, sample dies to the front of an STM platen (left) entails carousel rotation – with its attendant pressure transient – following cleavage so the sample surface and STM tip face one another. Fastening these dies to the back of the symmetric, STM platen (right) circumvents any need for carousel rotation, since cleavage surface and STM tip are naturally aligned.

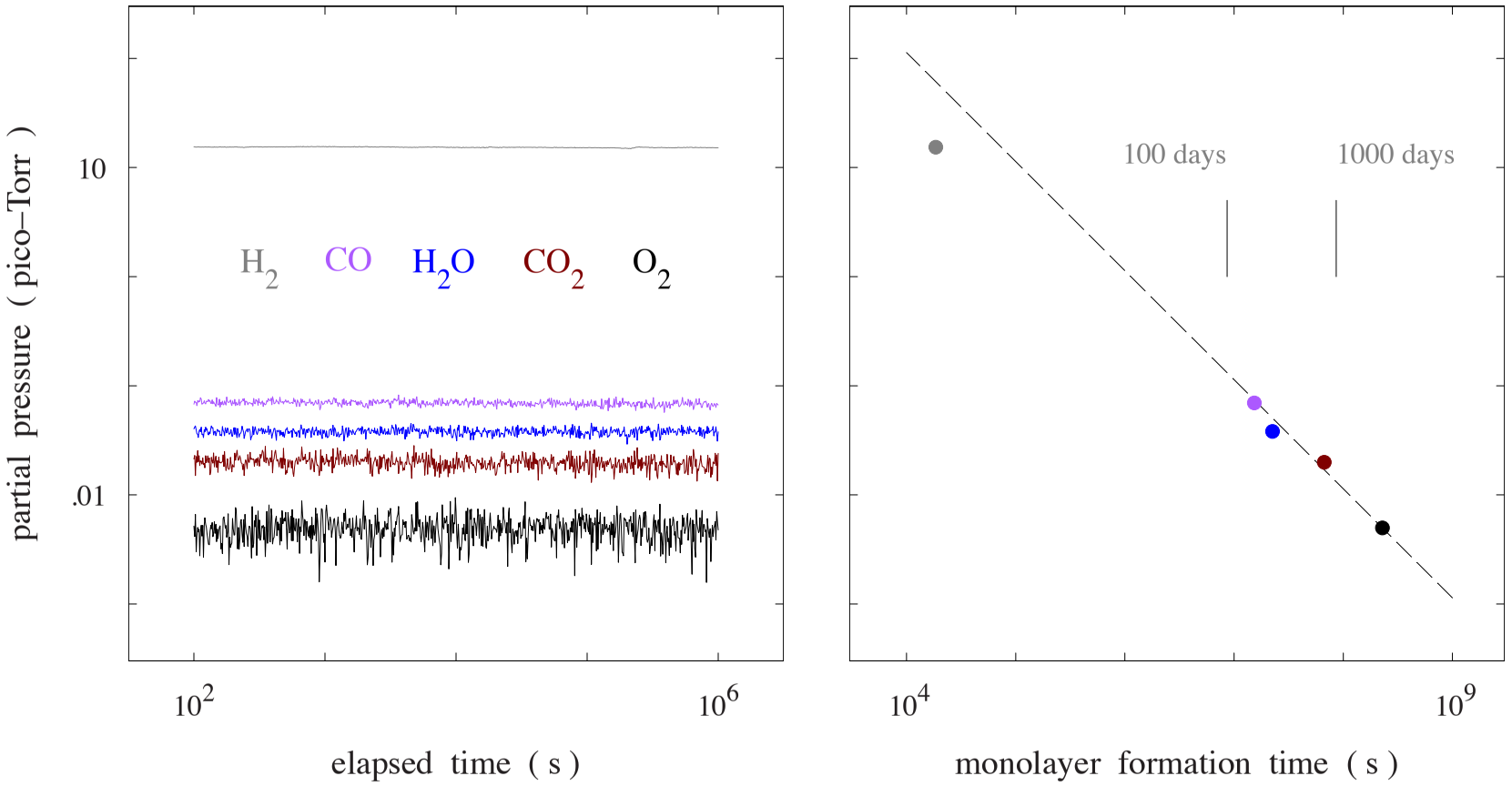


FIGURE 2.14. RGA partial–pressure analysis illustrating steady–state XHV conditions achieved with the upgraded UHV system (left) and associated (minimum) monolayer–formation times under ideal circumstances (right).

entirely eliminated, and only a handful of adsorbates can render an STM image (~20,000 reactive sites) useless for our purposes; thus, this (naive) estimate is overly optimistic by an order of magnitude or more; several days (up to a week) turns out to be a more realistic measure.

Although the painstaking improvements needed to reach this point took over two years to complete, they were well worth it, with the dramatic consequences of such a wholesale reconditioning for STM most readily appreciated by referring to the side–by–side, before–and–after comparison presented in Fig. 2.15. Our state–of–the–art chamber has reliably maintained this level of vacuum cleanliness for more than six years now, permitting its continuous, uninterrupted scientific use throughout that entire period.

Material Opportunities and Limitations

MQW heterostructures are fabricated by alternating semiconductor materials that possess the same crystal structure (i.e., zincblende) and similar lattice constants, but distinct band gaps, in a process often associated with the phrase “bandgap engineering”. The constituent layers throughout a MQW are often engineered to adopt the lattice of the substrate to eliminate any lattice–mismatch–induced stress during growth and thereby improve structural quality. By adjusting the In fractions of both InAlAs and InGaAs alloys it is feasible to simultaneously match their lattice constants to an InP substrate as indicated [44] in Fig. 2.16; this design is commonly referred as a lattice–matched (LM) structure. However, the wavelength agility achievable in QCLs with such designs is then limited by the constituent layer thicknesses.

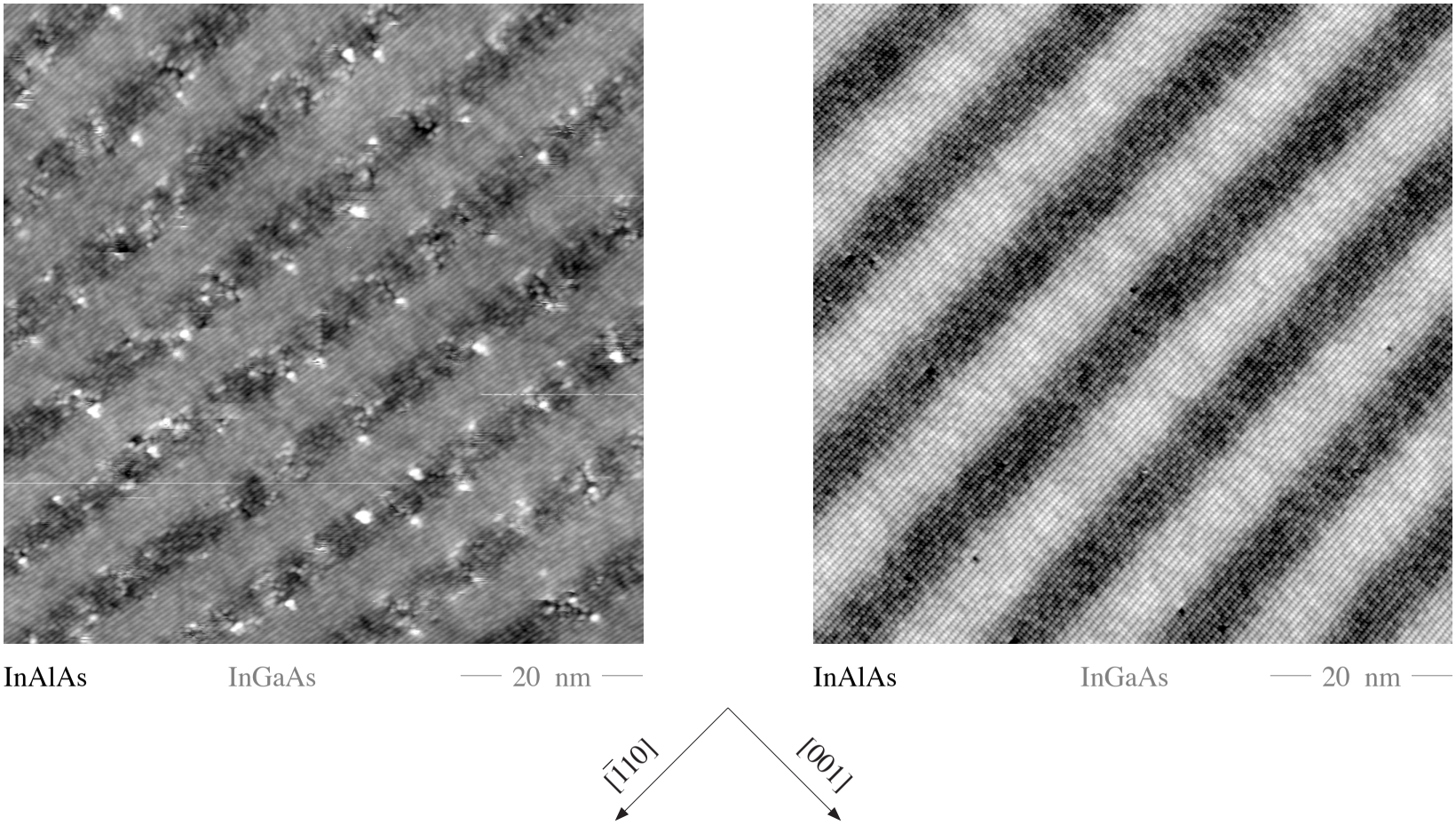


FIGURE 2.15. Typical STM images from a freshly–cleaved InAlAs / InGaAs SL following 24-hour exposure to the ambient environment of the original (left) and reconditioned (right) UHV system. All visible surface contamination has been effectively suppressed.

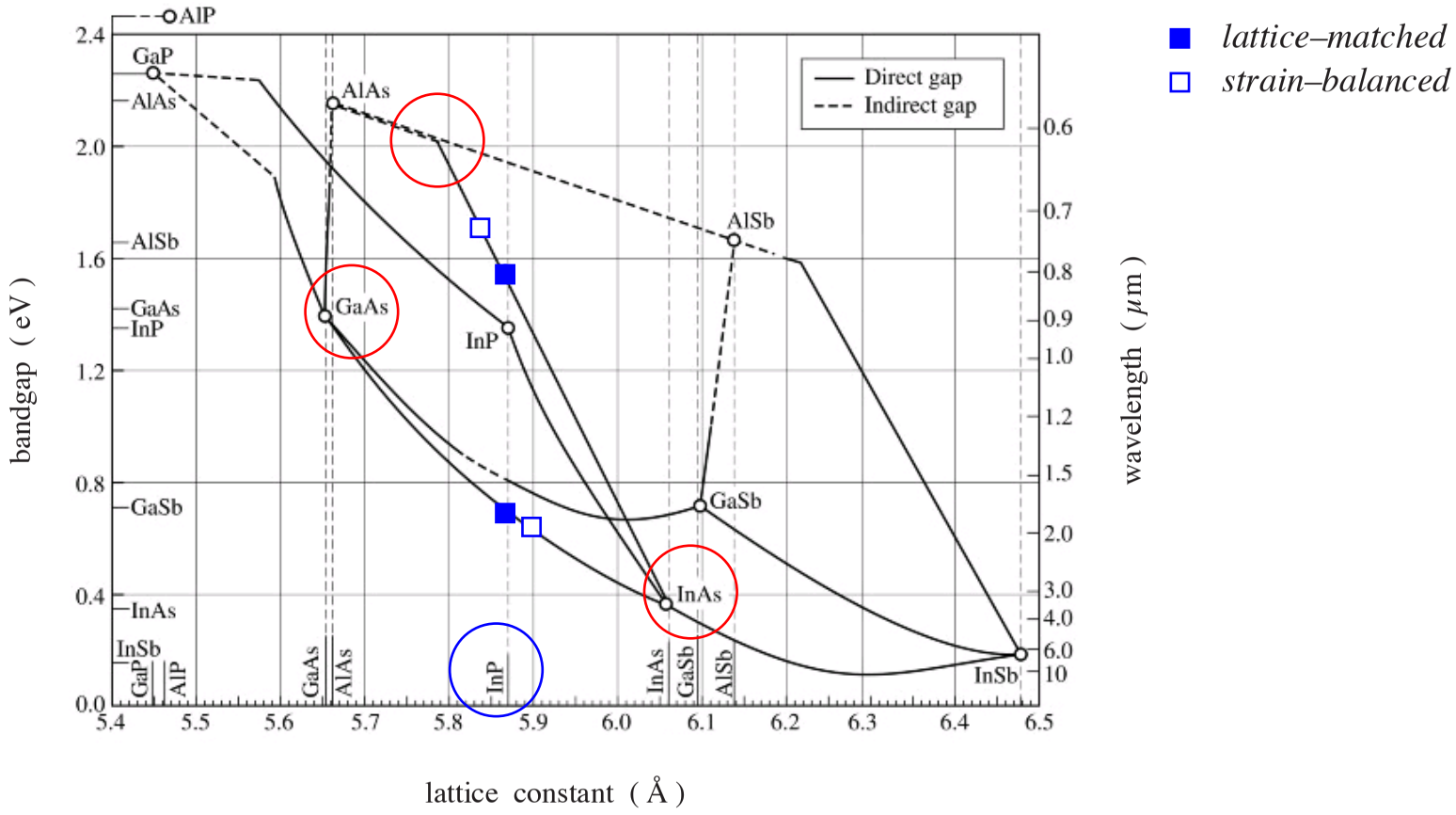


FIGURE 2.16. Bandgap energy and atomic spacing for common III / V semiconductor alloys. With careful control over the alloy composition of the constituent InAlAs and InGaAs ternaries it is possible to engineer these layers to either individually (solid squares) or cooperatively (open squares) adopt the lattice spacing of an InP substrate. Reprinted from the internet [44].

A more versatile approach is opened up by strain–balanced (SB) designs that tailor the well and barrier stoichiometries to produce deeper wells along with higher barriers, hence enabling shorter emission wavelengths. The deliberate variation in the alloy fractions of these designs imposes an important physical constraint since the constituent layers acquire slightly different, but potentially compensating, lattice constants than the substrate as indicated in Fig. 2.16. Strain–balanced structures alternate compressively–strained wells with tensilely–strained barriers by exercising precise control over *both* constituent layer thicknesses and alloy compositions during growth. In this way each constituent layer is pseudomorphically strained to the crystalline substrate yet, ideally, the net strain per superlattice period vanishes. Either of these epitaxial growths could, in principle, be used for STM studies, with their respective strengths and weaknesses discussed below.

Once substrate cleavage and vacuum cleanliness were successfully addressed, atomic–resolution images of the lattice–matched InAlAs / InGaAs superlattice quickly followed. This strain–free heterostructure was selected first (as opposed to its strain–balanced counterpart) to increase the likelihood of obtaining atomically flat cleavage throughout the epitaxial layers. The LM superlattice did indeed cleave perfectly with a very high probability of success, but another unanticipated problem was soon encountered. It turns out that STM images of LM superlattices do not possess the contrast necessary to clearly demarcate each constituent layer; and, as will become clear shortly, any qualitative analysis of interface roughness hinges on an ability to digitally delineate, and isolate, these layers from one another.

As previously explained, STM contrast in MQW heterostructures is influenced by two distinct factors: physical changes in the cleaved surface profile complemented by discontinuities in the energy bands. The latter contribution to the contrast in InAlAs / InGaAs LM and SB designs is illustrated in Fig. 2.17. It is clear from this diagram that the band offsets between InAlAs and InGaAs in the SB design are larger than those in the LM structure. This distinction between the two architectures points to the SB design as the superior candidate with respect to electronic layer contrast.

The contribution to the layer contrast arising from variations in the surface profile after cleavage requires a more careful consideration. Fig. 2.18 presents a simplified, two–dimensional view of the atomic arrangements in a LM design. Quantum wells (top–left) and quantum barriers (top–right) in these structures each naturally assume the atomic spacing of the substrate template in the constrained growth plane (horizontally in Fig. 2.18); they likewise assume the substrate lattice constant in the unrestricted growth direction (top–to–bottom in Fig. 2.18) as a direct consequence. Because the constituent layers in a LM design are not forced to deform during epitaxial growth there will be no strain relaxation post–cleavage and hence no deviations of the exposed surface contour to contribute to the layer contrast seen with STM. In other words, in ideal LM structures the only contribution to the layer contrast is entirely electronic.

We now revisit these arguments from the perspective of a SB design; Fig. 2.19 presents a simplified, two–dimensional view of this architecture. The quantum wells here (top–left) are designed with larger atomic spacing than the underlying substrate and

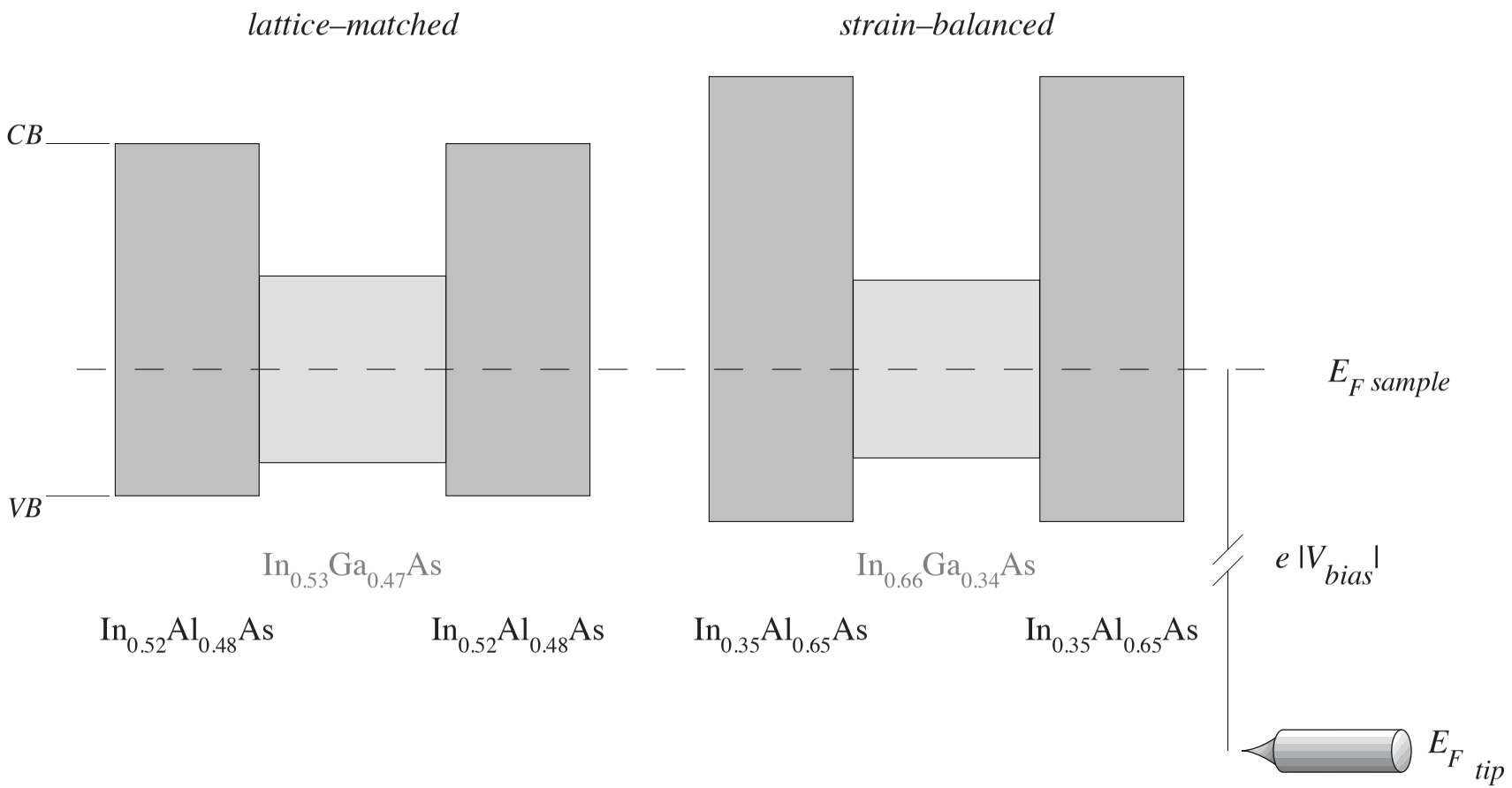


FIGURE 2.17. Energy band diagrams vs [001] displacement for LM (left) and SB (right) InGaAs / InAlAs MQWs. Under similar imaging conditions, the layer contrast expected with a LM design is weaker than with a SB design due to the smaller valance–band offset between wells and barriers. Vertical and horizontal axes (energy and distance along the growth direction) are both to scale. The STM tip is illustrated as if tunneling occurs at the (001) growth surface to emphasize the relative energy alignments; in reality, the tip is rastered left to right across this diagram in a perpendicular {110} surface.

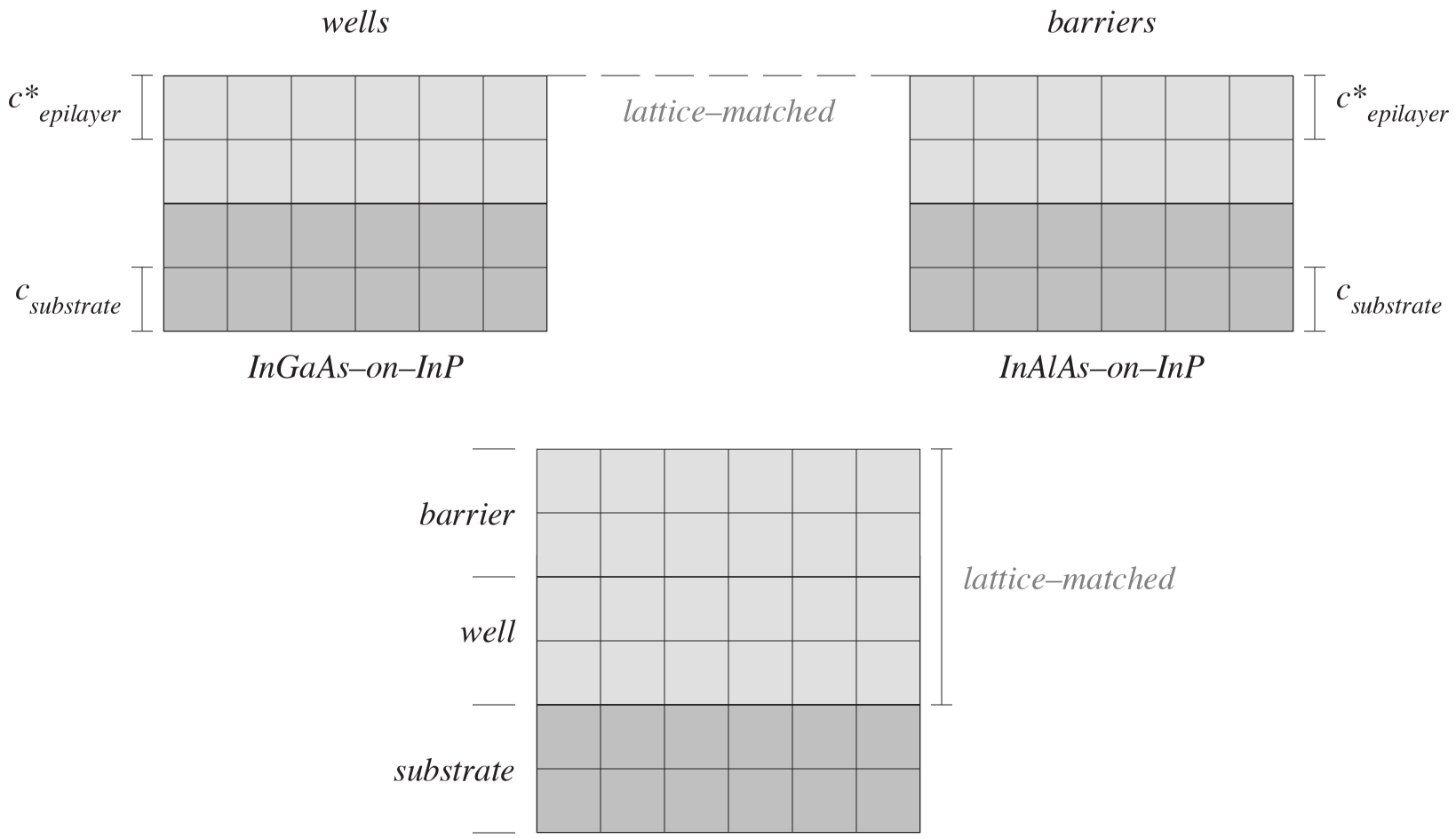


FIGURE 2.18. 2–D schematic showing InGaAs (top left) and InAlAs (top right) lattice matched to InP. Since the constituent layers in this design naturally adopt the atomic spacing of the underlying substrate, both in–plane (horizontally) and in the growth direction (vertically) there is no bulk relaxation following cleavage (bottom). The STM layer contrast in a lattice–matched architecture is (therefore) entirely electronic.

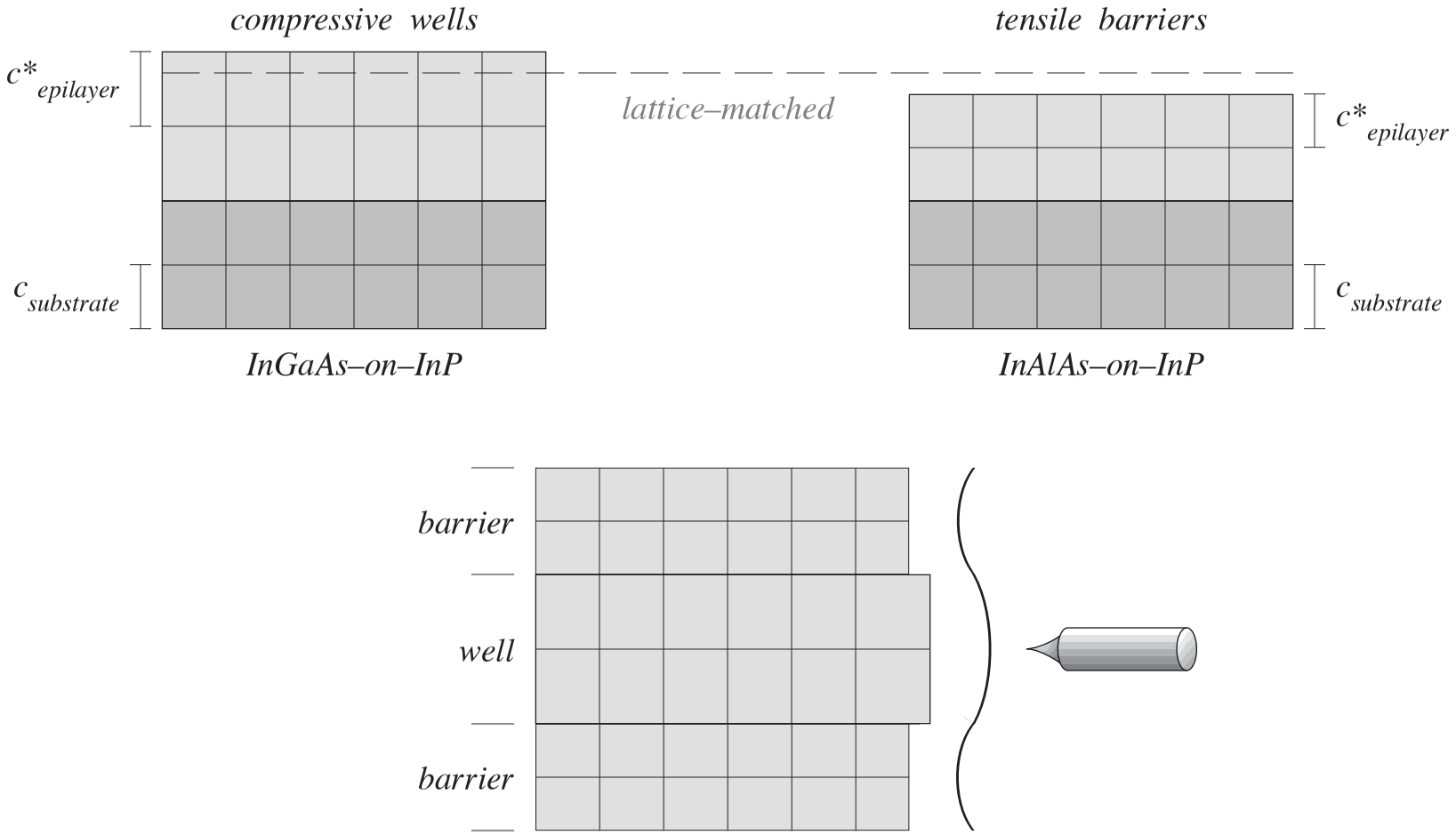


FIGURE 2.19. 2–D schematic showing compressively–strained InGaAs (top left) and tensilely–strained InAlAs (top right), each coherently matched to InP. The constituent layers in this design must deform to adopt the atomic spacing of the underlying substrate in the constrained growth plane (horizontally), but are free to relax in the growth direction (vertically). The in–plane strain accumulated during growth is released once the constituent layers are free to return to their natural geometry along a cleavage exposed facet (bottom). This relaxation contributes to the STM contrast in strain–balanced architectures.

are thus forced to *compress* in the constrained growth plane. The elastic strain associated with this deformation is free to relax towards the unrestricted growth direction and therefore extend past the nominal substrate atomic spacing (dashed line). This crystalline deformation is carefully compensated with quantum barriers (top–right) that are designed with smaller atomic arrangement than the substrate and must therefore *expand* their bond lengths in the growth plane during epitaxial deposition to perfectly occupy the underlying surface template. The strain relaxation in this case has the opposite effect so that quantum barriers adopt a lattice constant in the growth direction smaller than that of the substrate (dashed line). Precise control over the alloy composition and layer thicknesses of both constituent materials makes pseudomorphic growth of strain–balanced multilayers with zero net strain in the growth direction possible. Strained layers, once exposed in cross section through cleavage, will relax back to their natural geometry along the newly created free surface. Compressively–strained layers are pushed outward, and tensilely–strained layers drawn inward (see Fig. 2.19, bottom) partially relieving the elastic stress accumulated during growth of each constituent [45, 46]. With compressive wells and tensile barriers, as is the case in our studies here, electronic and topographic factors will *constructively combine* to produce the total layer contrast seen in filled–state STM images.

Since individual layer identification in LM heterostructures does not appear promising (Fig. 2.20, left) our STM focus from this point forward will be on strain–balanced InAlAs / InGaAs superlattices (Fig. 2.20, right). This design – fabricated by the same practitioner under similar growth conditions – possesses the same layer geometry

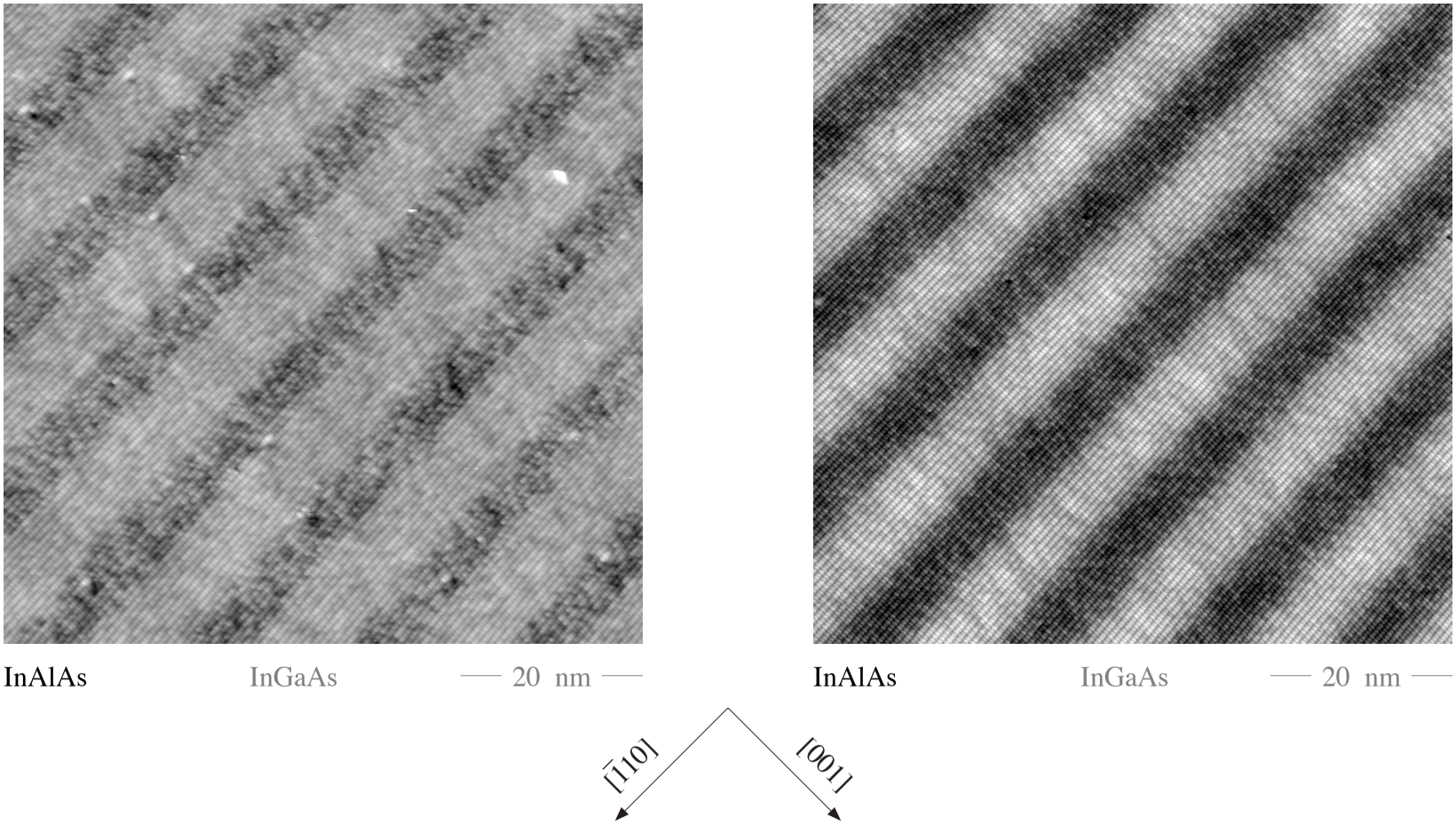


FIGURE 2.20. Representative, atomic–resolution STM images of the arsenic sublattice contrasting nominally identical InAlAs / InGaAs lattice–matched (left) and strain–balanced (right) superlattices. Under similar imaging conditions, the strain–balanced structure exhibits significantly stronger layer contrast than the lattice–matched design.

and the same elements as its lattice–matched counterpart, differing only in alloy composition. Even though, this "small" modification makes cleavage more challenging, the resulting increase in layer contrast proves essential for a clear identification, and delineation of wells versus barriers. Grey–level histograms[2] from the two designs are compared in Fig. 2.21; it is evident that the SB heterostructure manifests a definite bimodal distribution separating well– from barrier–like intensities.

Instrumental Non–idealities

Two major steps are required to accomplish our research objectives: first, heterojunctions must be clearly identified and isolated; second, the interface roughness at each of these heterojunctions must be quantified and described statistically. We address these challenges in turn, beginning here with how to accurately determine the length scales involved in the description of interface roughness and postpone until Chapter III a discussion of how the interfaces themselves are identified.

Two significant analysis obstacles stand in the way of our research objectives. The piezoelectric raster mechanism in STM suffers from several well–known non–idealities that make accurate length comparisons problematic. As part of this work, we developed new and robust data acquisition protocols that transform this inherently “bad ruler” into a reliable atomic “yardstick” suitable for interface roughness measurements, thereby mitigating these long–recognized deficiencies. I am grateful to several current,

[2] Histograms shown here are normalized to unit area; this means that vertical and horizontal ranges are not the same for both desings. LM grey–level bandwidth is about half that of the SB, which means its intensity is about twice that of the SB design

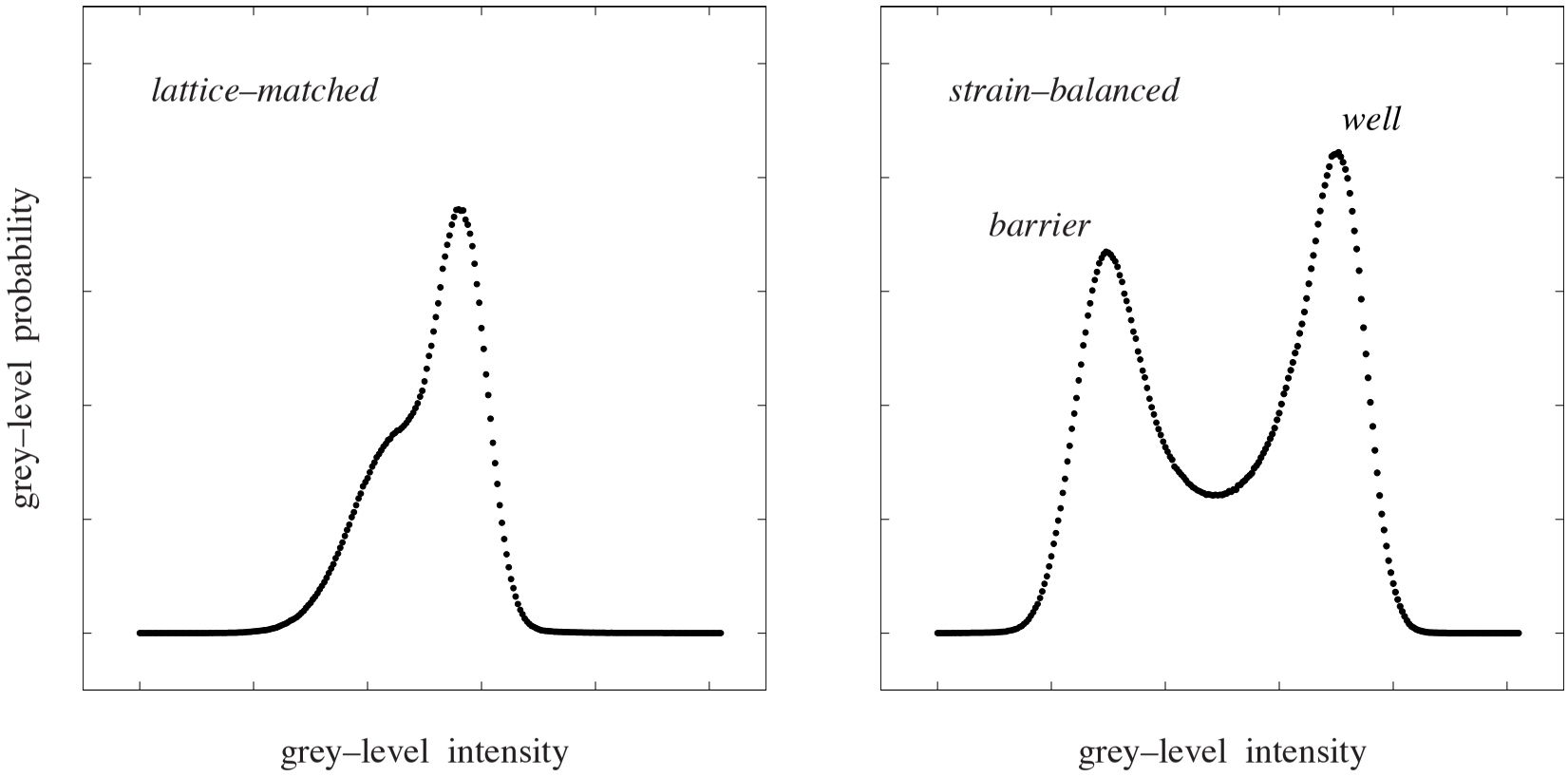


FIGURE 2.21. Survey–averaged, grey–level histograms from lattice–matched (left) and strain–balanced (right) InAlAs / InGaAs superlattices. Under similar imaging conditions, the strain–balanced design gives rise to a distinctly bimodal distribution reflecting individual well and barrier components; the corresponding lattice–matched histogram, on the other hand, does not readily permit an equally–clear division. Atomic corrugation has been suppressed in the images used here to emphasize the superlattice contribution to the (composite) histogram.

and past, members of the STM laboratory who played important parts in making this advance possible.

Our approach is best illustrated by an easily understood example: measurement of a *local* SL periodicity with STM. Determination of atomic spacings with STM is most conveniently and accurately done by performing a two–dimensional (2–D) discrete Fourier transformation (DFT) of real–space images to the complementary reciprocal–space domain. This transition from real to reciprocal space is most transparently understood if we consider, first, a simple periodic structure such as an InP substrate. Fig. 2.22, top, shows the in–plane phosphorus sublattice imaged with STM following cleavage. The corresponding power spectrum, or reciprocal–space map (Fig. 2.22, bottom), possesses distinct spots localized about the [001] and [110] reciprocal–lattice vectors defining the two–dimensional surface mesh. Referencing all length measurements within a given STM image to these *locally* determined reciprocal–lattice vectors provides a ruler automatically calibrated in units of the underlying crystal structure's lattice constant.

Consider, next, a strain–balanced InGaAs / InAlAs superlattice. As motivated in connection with Fig. 2.21, this structure possesses strong contrast between the constituent layers along with a uniformly periodic well and barrier geometry. The corresponding image (Fig. 2.23, top) here reveals the in–plane arsenic sublattice, and the accompanying power spectrum (Fig. 2.23, bottom) again possesses distinct spots at the [001] and [110] reciprocal–lattice vectors. The new feature here, however, is the set of zone–center and [001]–convolved satellite peaks introduced by the superlattice

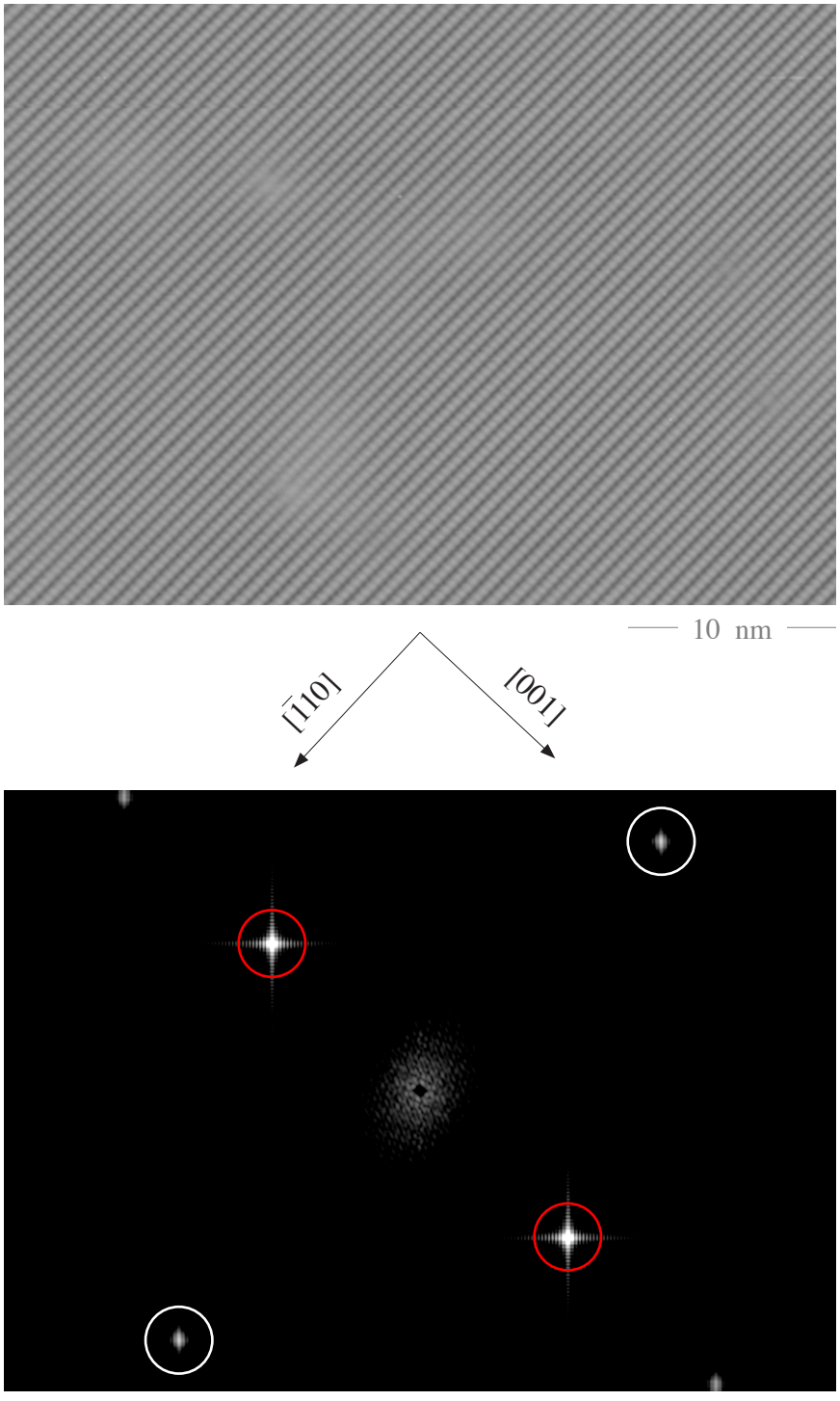


FIGURE 2.22. Atomic–resolution STM image (top) of the phosphorus sublattice over an InP substrate, together with its corresponding reciprocal–space power spectrum (bottom). Bright spots oriented along the growth (red) and in–plane directions (white) denote the two–dimensional surface mesh.

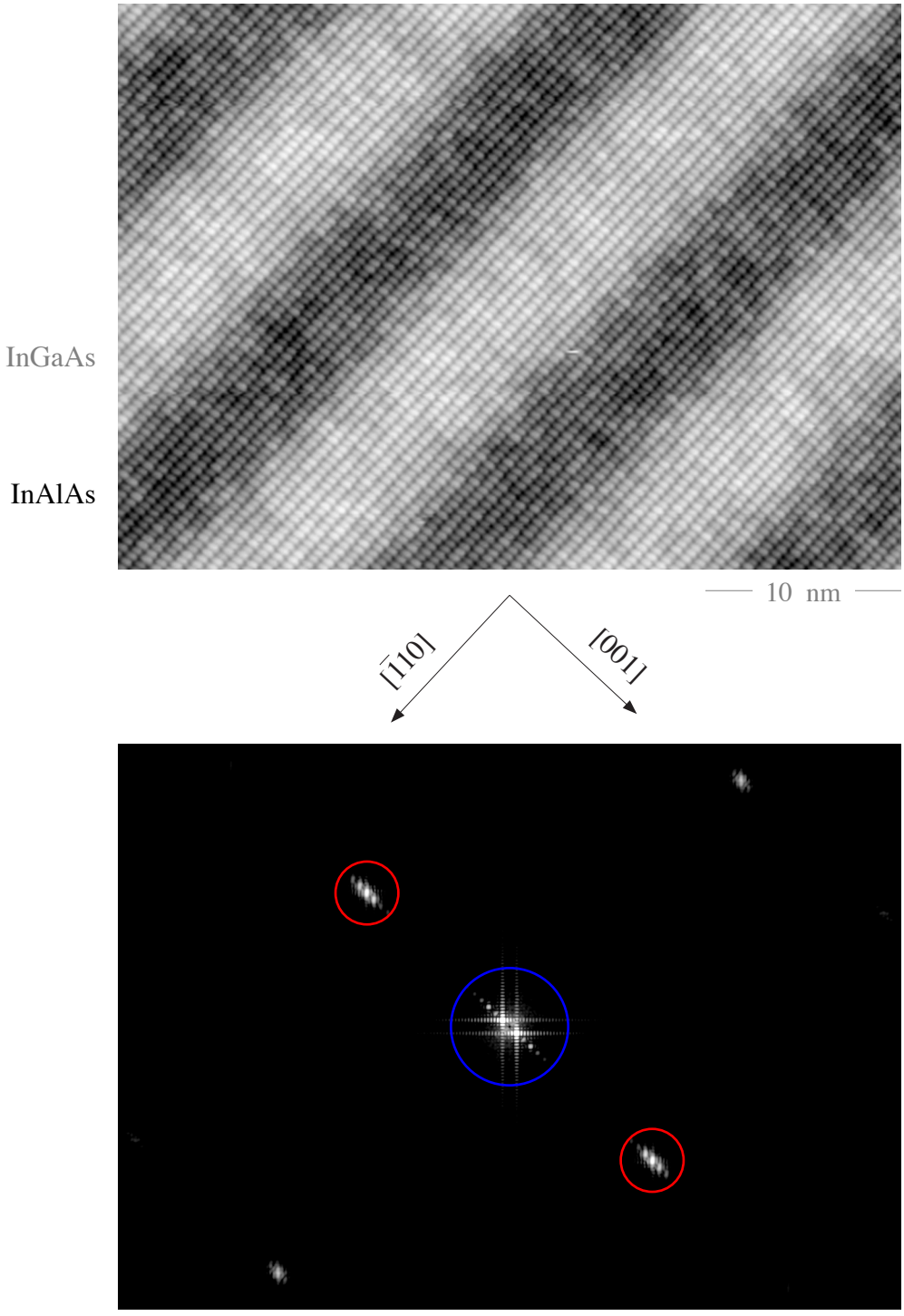


FIGURE 2.23. Atomic–resolution STM image (top) of the arsenic sublattice over a SB InGaAs / InAlAs superlattice, together with its corresponding reciprocal–space power spectrum (bottom). Additional peaks located near the zone center (blue) and in the vicinity of the [001] reciprocal–lattice vector (red) arise from the SL periodicity.

periodicity superimposed on the [001] atomic spacing.

The [001]–convolved satellite peaks in Fig. 2.23 are STM analogs of the (004) satellites routinely observed with conventional x–ray rocking curves. The distinct advantage offered by STM is its ability to restrict its focus to a specific subset of epitaxial layers rather than the entire vertical stack (Fig. 1.6). Under appropriate circumstances, then, a superlattice periodicity, measured in units of the local [001] reciprocal–lattice vector, may be obtained from the slope of the corresponding Bragg–like plot. As we show below, this approach places certain STM measurements on nearly an equivalent footing with x–ray diffraction in terms of absolute accuracy; a point apparently overlooked by the STM and TEM communities.

In this regard, it is important to appreciate that the conversion from real to reciprocal space reveals not only the underlying atomic periodicity of a sample, but also the deviations from it introduced by a non–ideal raster mechanism. For example, the 2–D DFT of a perfectly–periodic substrate should produce delta–function–like spots in reciprocal space; any broadening in excess of that anticipated due to finite size effects must be attributed to image distortion. Splitting of reciprocal–lattice spots, in one or both scan directions, is another potentially related artifact due to piezo nonlinearities.

The typical distortion processes affecting STM images are either temporal[3] or instrumental [47, 48] in nature, both of which can be managed to some extent. Temporal distortion is frequently attributed to poor practice in STM data acquisition, and thus reasonably remedied. The sample stage used to (coarsely) maneuver a region of interest

[3] After coarse movement the STM–sample–stage has a tendency to continue in motion because of inertia; it takes a few hours for the STM mechanism to satisfactorily settle.

directly beneath the STM tip serves as an appropriate example. It is well–known that the piezoelectric ceramics actuating this stage display inertia [49] and have a tendency to continue in motion along the same direction for a considerable period of time as illustrated in the simplified schematic in Fig. 2.24. This problem is simply addressed by allowing sufficient time for the sample positioning mechanism to fully relax prior to image acquisition, thus eliminating any induced drift of the scan frame over the surface.

Instrumental distortion, on the other hand, is more difficult to identify and control. It arises from two undesirable features of a rectilinear tripod scanner illustrated in Fig. 2.25: piezo creep [47] (left) and piezo nonlinearity [48] (right). The nonlinear response of a piezo–ceramic to the linearly–varying scan voltage applied across it produces a corresponding drift, or "chirp", in the spatial frequencies, or reciprocal–lattice "spots", picked up by the DFT. Under certain circumstances, these spots not only broaden but also bifurcate into distinct points – the aforementioned splitting – making identification of a unique reciprocal–lattice vector impossible. One workaround for this unavoidable complication is to carefully constrain the real–space region over which the DFT is performed. By choosing a limited area of 400 Å by 400 Å, centered horizontally and near the top of each image where the *x*– and *y*–piezos are most nearly linear, suitably sharp, well–defined reciprocal–lattice vectors are routinely obtained.

Piezo creep is a sluggish (as opposed to instantaneous) change in piezo elongation (or contraction) following any change in the applied electric field. The geometric symmetry of an ideal tripod scanner is broken by the time asymmetry introduced by fast– and slow–scan directions aligned with *x*– and *y*–piezos respectively.

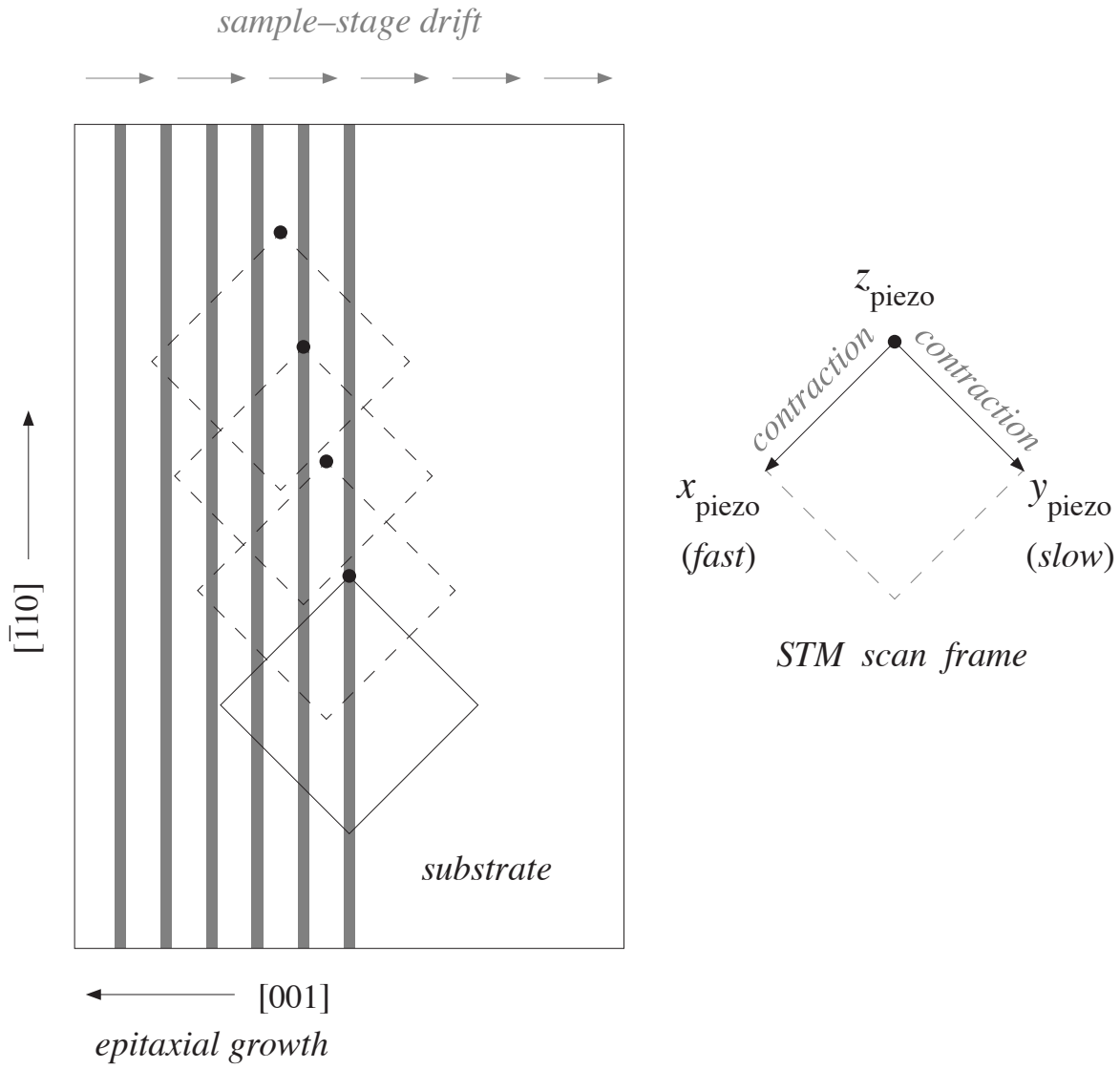


FIGURE 2.24. Schematic diagram illustrating apparent drift of the STM scan frame due to piezo–inertia of the sample stage (left). This stage permits sample movement either left (in the growth direction), or right (opposite growth direction) with respect to the tip, as illustrated here. The induced drift typically requires several hours to settle. Also indicated in this diagram, is the STM scan frame (right), where contraction of the *x*– and *y*–piezos produce positive displacements. Cleavage plane coincides with plane of paper. Growth direction is right to left.

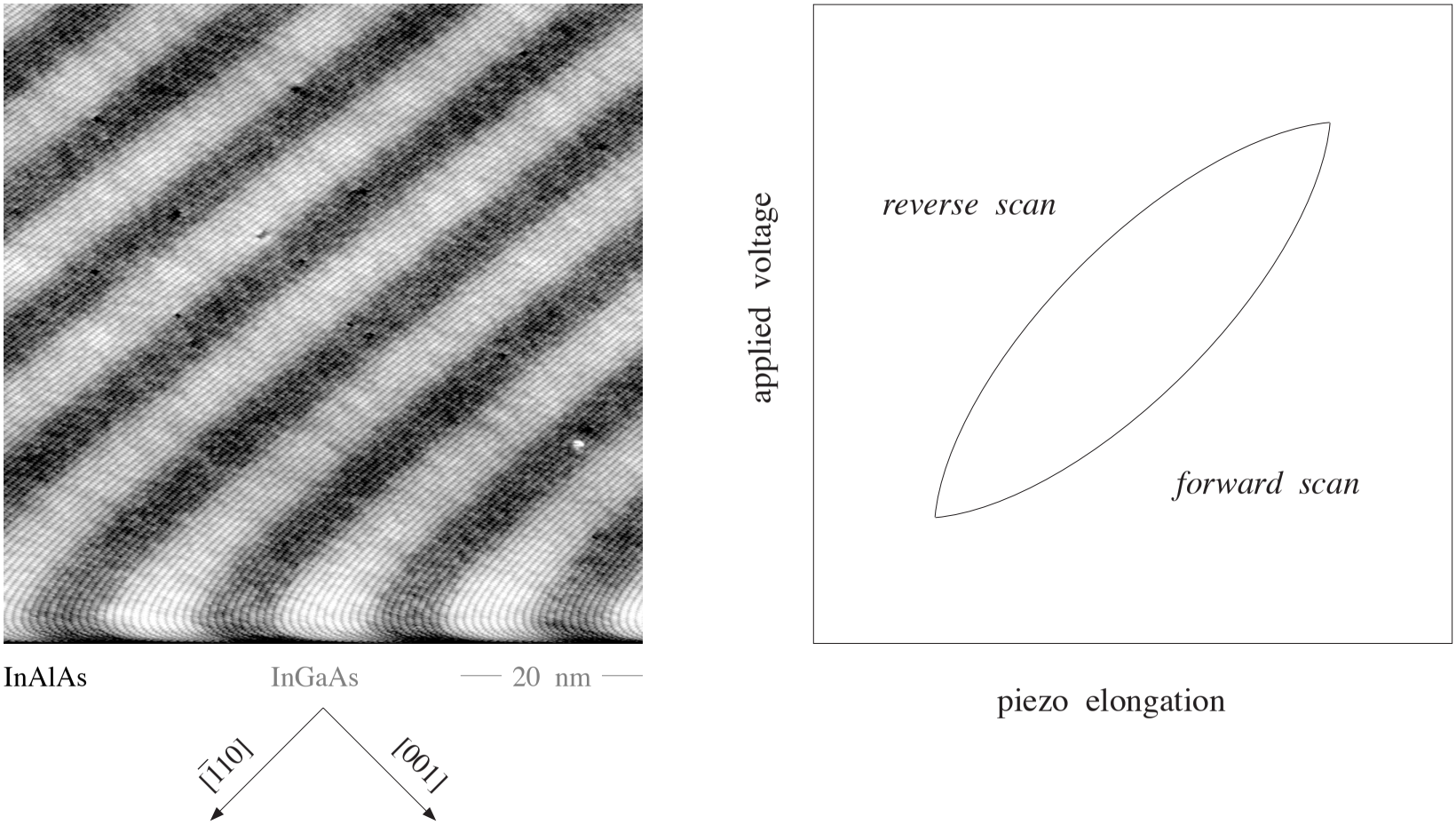


FIGURE 2.25. Typical non–idealities in piezo–scanners include piezo–creep (left) and hysteresis (right). Piezo–creep is visible here as whiplash in the bottom of the image; a subtler, but nevertheless undesirable symptom of piezo–creep also occurs throughout the image if the navigation route *y*–offsets are inadvertently taken to be parallel to the slow–scan direction. Piezo non–linearity (hysteresis) is on the other hand unavoidable, a workaround to ensure reciprocal–lattice vectors are sharp and well–defined is to use a restricted area (400 Å by 400 Å) where piezos are most linear in our current setup.

This asymmetry causes piezo–creep to become a particularly important issue in the slow–scan direction. The significance of this point becomes clear once navigation over the cleavage–exposed surface is fully explained. As already noted, the coarse positioning stage is not suited for this purpose since its settling time is too long. Instead, adjustable dc offsets directly applied to the *x*– and *y*–piezos may be used to maneuver the tip anywhere within a one–micron–square window as illustrated in Fig. 2.26. These offsets are independent of the ac voltages used to sweep thru the (much smaller) image frame, but the positive, slow–scan direction is (by default) aligned with the positive *y*–offset direction. Piezo creep in the *y*–direction is especially pronounced whenever the *y*–offset and slow–scan increment are (incorrectly) chosen parallel to one another, and correspondingly attenuated whenever the two are (correctly) taken anti–parallel.

The fast–scan direction is (comparatively) immune to this distinction because the corresponding scan increment is bipolar: the fast–scan ceramic first contracts (forward) and then expands (reverse) with each line of the raster. The slow–scan increment that follows each *x*–raster, on the other hand, is unipolar, so the slow–scan ceramic always contracts.

Standardized Navigation Routes

These practical considerations distinguish two fundamentally different navigation routes, denoted as minimally– and maximally–distorting, respectively, in Fig. 2.26. Since negative (anti–parallel) navigation offsets of the *y*–piezo are needed to keep creep to a minimum, the sign of the corresponding *x*–piezo offset is entirely determined by

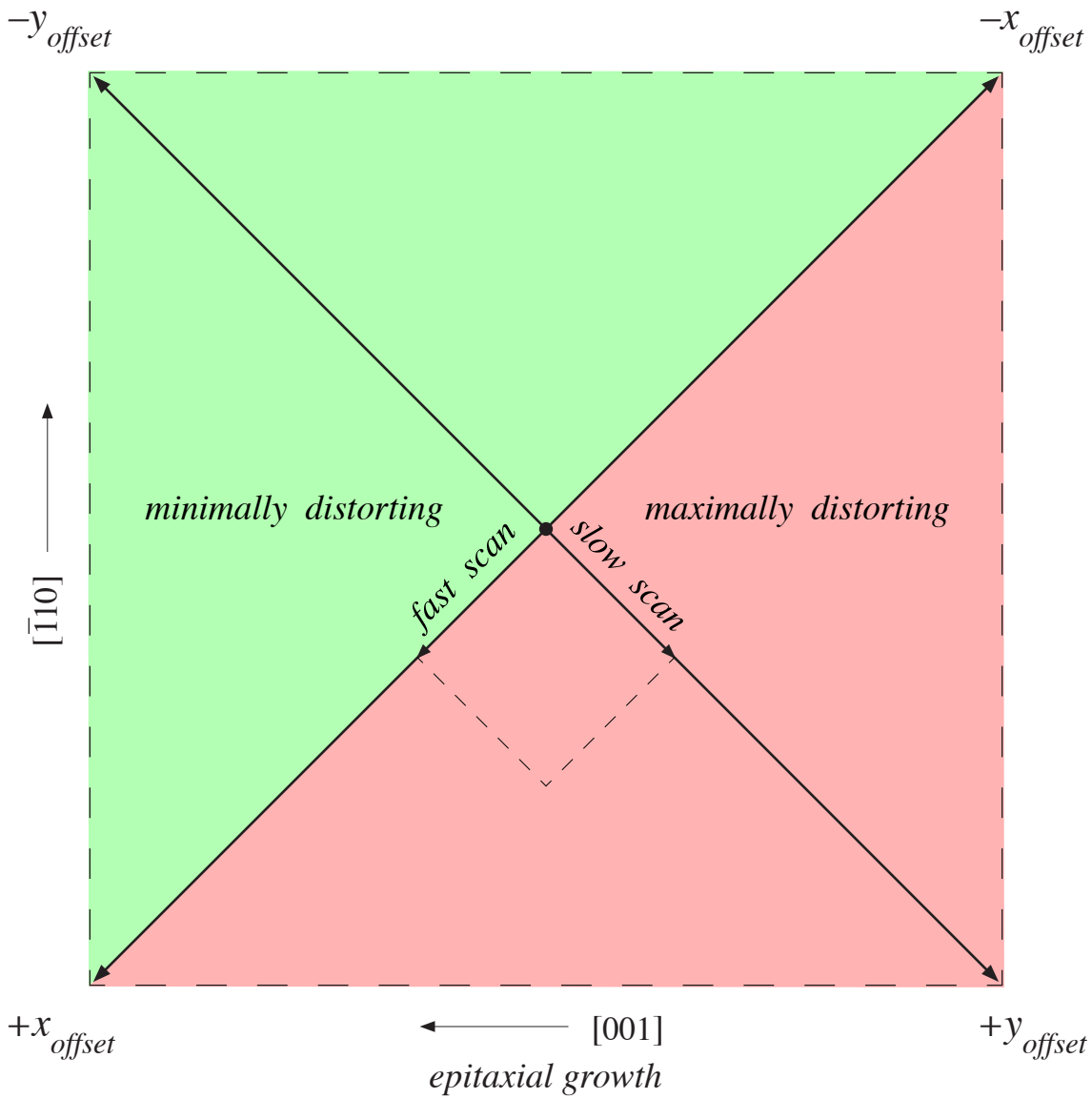


FIGURE 2.26. STM image frame located at the origin indicates fast– and slow–scan directions oriented along x and y piezo axes, respectively. This image frame can be precisely positioned anywhere within the one–micron–squared maximum area available to the tripod scanner with fine adjustment of the x– and y–piezo offsets. Minimally distorting zone corresponds to navigation routes that avoid positive increments of the y–piezo offset. Growth direction is right to left.

vector geometry. For lateral surveys that focus on a fixed subset of periods along the [–110] direction, this offset will be necessarily positive; for transverse surveys that explore the structure from beginning to end in the [001] growth direction, this offset, like the one applied to the y–piezo, will be negative.

Examples of device–scale STM surveys acquired with these standardized routes are presented in Fig 2.27 for an InAlAs / InGaAs strain–balanced superlattice. Two of the three surveys shown in Fig. 2.27 are *lateral* surveys acquired on the cleavage plane along a particular subset of periods; one survey was acquired close to the beginning of the growth and the other near the conclusion; the remaining, *transverse*, survey was acquired by navigating across the entire epitaxial structure in the growth direction. The arrows (shown in Fig. 2.27) indicate the standardized navigation routes followed.

Initial coarse positioning with the STM sample stage additionally relies on easily identifiable markers signaling the beginning and conclusion of epilayer growth. Referring back to the SB superlattice surveys in Fig. 2.27, the onset of superlattice growth is identified by the unmistakable transition from a homogeneous InP buffer layer (and transitional InGaAlAs quaternary layer) to a modulated InAlAs / InGaAs heterostructure. Termination of the superlattice is similarly identified by the reciprocal transition from modulated InAlAs / InGaAs to a homogeneous InP capping layer (also indicated in Fig. 2.27) whose sole purpose is to facilitate STM navigation. Once a general location has been so identified, the STM tip can be repositioned over a particular area of interest with precise adjustment of the scan–frame offset.

The novelty of our data acquisition protocol resides in the meticulous and

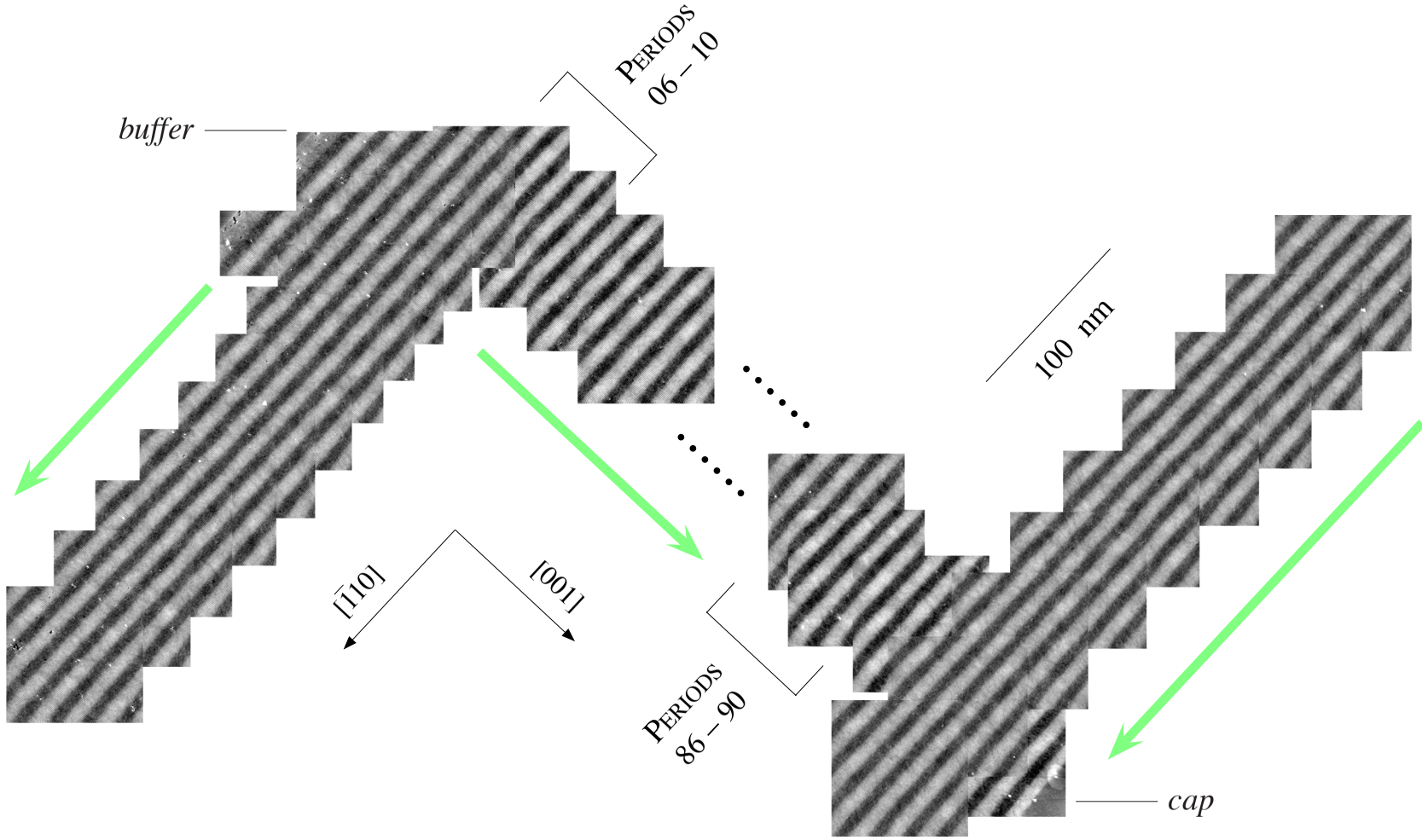


FIGURE 2.27. Device–scale, atomic–resolution surveys acquired over a SB InAlAs / InGaAs superlattice. The lateral survey on the left was acquired near the beginning of the growth as indicated by the InP buffer layer, whereas the lateral survey on the right was acquired towards the conclusion of the growth similarly indicated by the InP capping layer. The two lateral surveys are jointed by a transverse survey obtained across the one–hundred repeats. Green arrows indicate the direction each survey was acquired. Growth direction is from top–left to bottom–right.

accurate navigation command exercised during micron–long excursions in the cleavage plane along any given interface. Heterojunctions selected for examination remain fixed with respect to the imaging field of view, shifting by no more than two monolayers in the growth direction across an entire survey. There are two reasons why this is especially helpful. First, that the location of each heterojunction is fixed with respect to the scan frame provides important confirmation the STM instrumentation has settled, and that all reciprocal–lattice vectors will be sharp and (reasonably) stable over time. The second reason is related to data analysis, since it facilitates comparisons between independent surveys that target nominally the same area, as we will see shortly.

STM Superlattice Period Determination

The asymmetric susceptibility of *x*– and *y*–scan components to piezo creep discussed above suggests any attempt to use the magnitude of the [001] reciprocal–lattice vector as a *precision* distance metric is likely to be ill–fated, our carefully–considered navigation protocols notwithstanding. This was indeed found to be the case experimentally.

A more refined approach that explicitly recognizes this asymmetry turns, instead, to the separate *x*– and *y*–components of the atomic reciprocal–lattice vector (and, by extension, the analogous *x*– and *y*– components of the corresponding superlattice wave vector) obtained from DFTs of the real–space images. Fig. 2.28 summarizes these components for the particular lateral survey illustrated in Fig. 2.29 (left), but the results illustrated there are typical in lateral surveys. Two important conclusions immediately

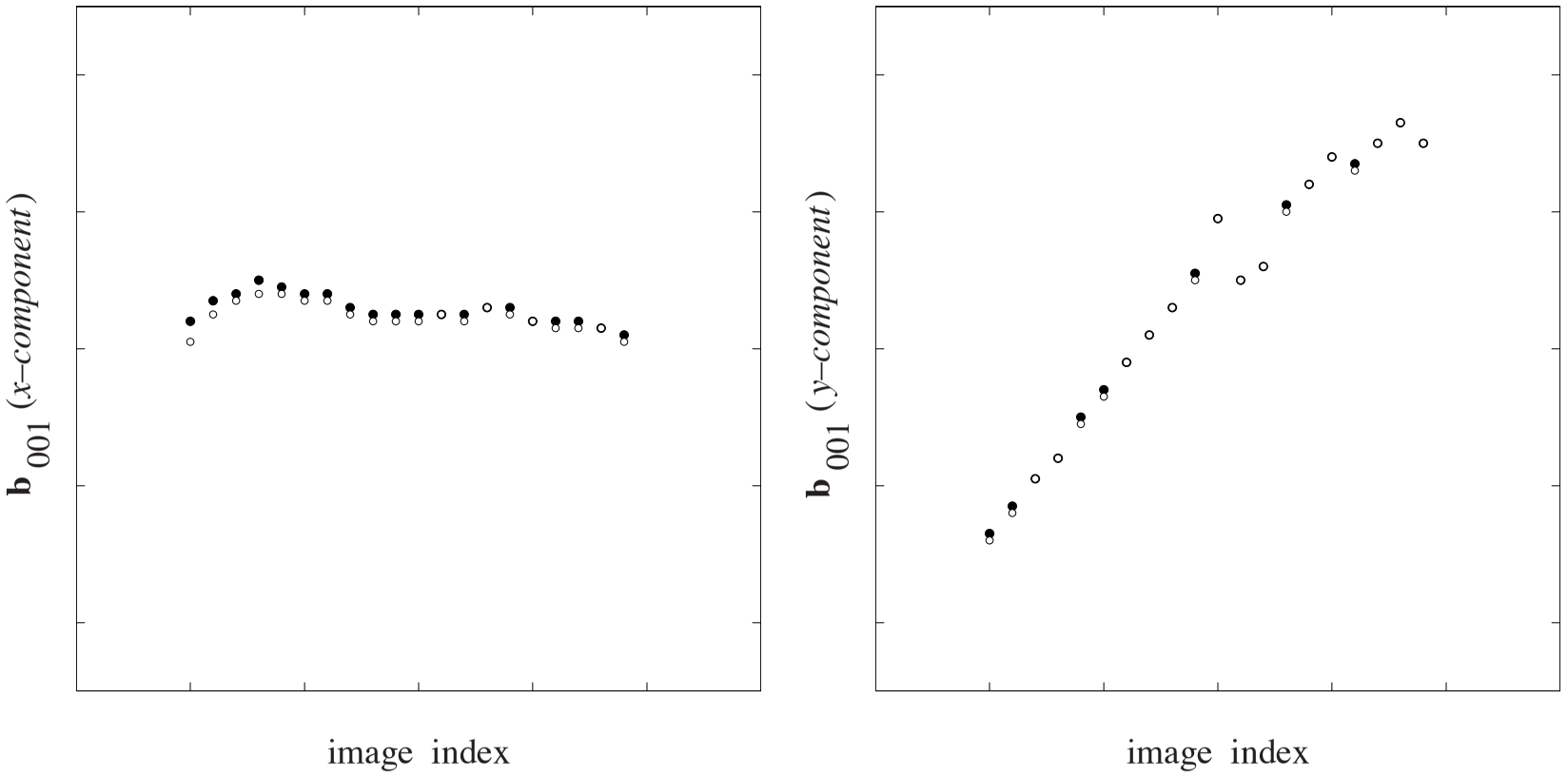


FIGURE 2.28. Representative behavior of the [001] reciprocal–lattice vector *x*– (left) and *y*– (right) components during a lateral survey. Vertical range (with suppressed zero) is the same for both components clearly indicating the larger drift in *y* throughout the survey. Full range for *x*– and y–component is ~ 0.6% and 4% of respective mean value. Horizontal axis is image number, which is the same in both plots. Piezo hysteresis notwithstanding, the reciprocal–lattice vectors from forward– (closed circles) and reverse– (open circles) scans agree remarkably well (better than a part–per–thousand).

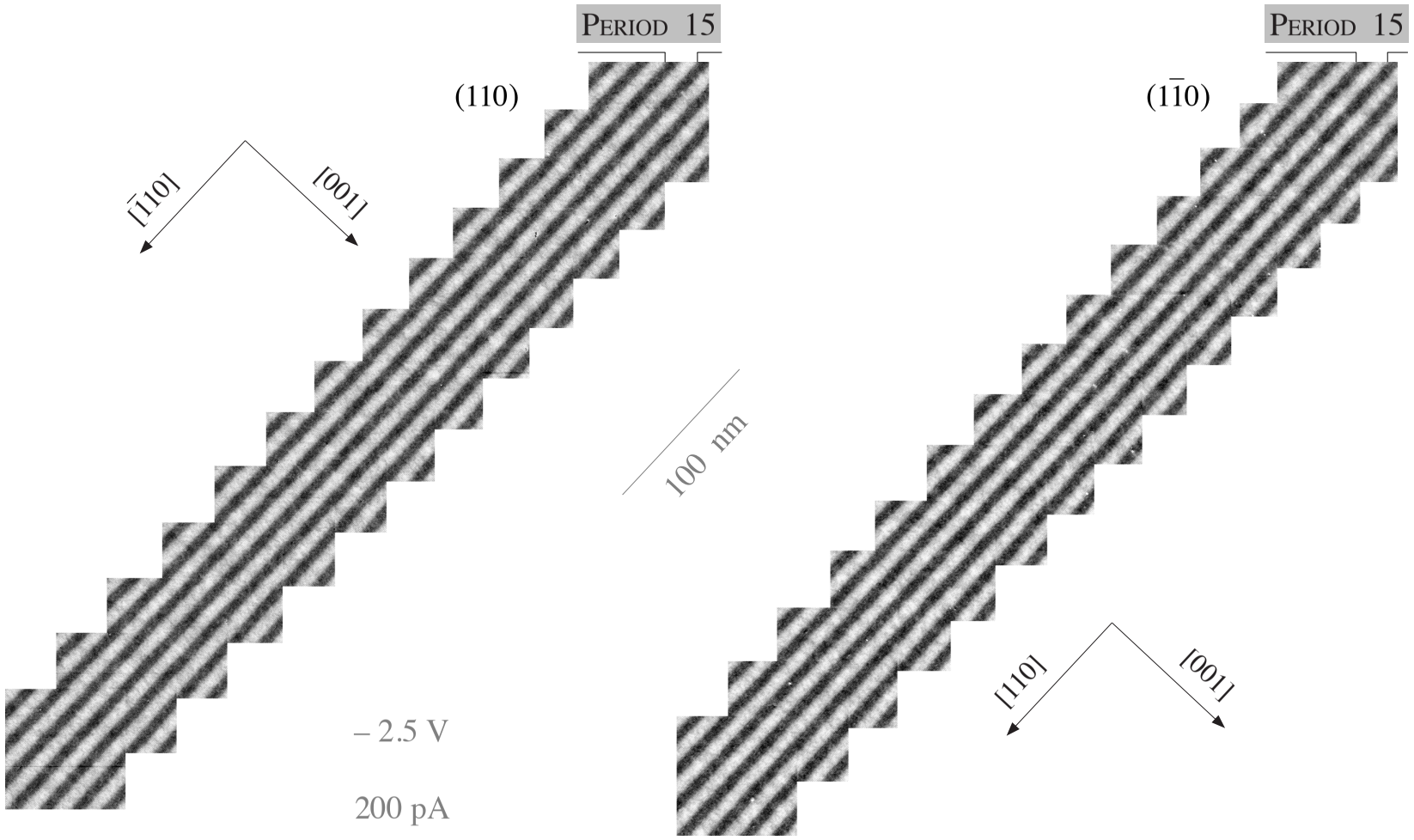


FIGURE 2.29. Device–scale, atomic–resolution surveys acquired over identical periods of a SB InAlAs / InGaAs superlattice in orthogonal (110) and (1-10) cross sections. The respective quantum well and barrier components of the superlattice are identified in Fig. 2.23, top. Growth direction is from top–left to bottom–right in both surveys.

jump out from this figure: first, the *x*–component of the [001] reciprocal–lattice vector is manifestly more stable (as a function of image index, or equivalently time) than the *y*–component; this is unsurprising in light of piezo creep affecting the *y*–component more strongly. Second, forward– and reverse–scans agree remarkably well with one another, any difference between the two remaining small (better than a part–per–thousand); this is perhaps surprising in light of piezo hysteresis, but also useful in terms of the data pooling.

The "real–time" feedback provided thru image–by–image inspection of a reciprocal–space power spectrum obviously plays a decisive part in vetting the navigation routes that will minimize instrumental distortion. What we explain next is that under survey conditions known to yield sharp, well–defined reciprocal–lattice spots, this power spectrum also affords a breathtakingly accurate and reproducible measurement of the local superlattice period.

Minimally–distorting, atomic–resolution surveys were conducted over (110) and (1–10) cross sections of the SB superlattice described in Fig. 1.6. Clearly–resolved atomic–scale features were used as registry guideposts, first, for precise navigationduring data acquisition and, later, for the careful alignment of successive scan frames needed to construct a survey mosaic from individual images, in each case. The mosaics from two such surveys acquired over identical periods in orthogonal cross sections are assembled in Fig. 2.29. These data constitute a canonical image set we will return to frequently in succeeding chapters, but they are used here solely to compare STM and HRXRD measurements of the superlattice period. That comparison serves as a

benchmark for our success in transforming an inherently problematic length–scale measurement tool — STM — into a suitably accurate one for quantifying interface fluctuations.

Figure 2.30 shows an x–ray rocking curve, taken from the center of the substrate wafer, along with the ensuing Bragg plot of superlattice peak angles versus satellite order [50]. From the (inverse) slope of the best–fit line to these data, one finds an average period of 40.71 ± 0.01 ML for the entire stack of 100 repeats, a value approximately 4% larger than the targeted 39 ML design.

The STM survey shown in Fig. 2.29 (right) was acquired over a location likewise close to the center of the substrate wafer (Fig. 2.3) but straddling only a handful of the 100 superlattice periods otherwise sampled with x–ray diffraction. An [001] section through the survey–averaged power spectrum (Fig. 2.31 top) reveals [001]–convolved satellite peaks analogous to those observed with HRXRD (Fig. 2.30 top), whose separation as a function of peak order (Fig. 2.31 bottom), when similarly analyzed (Fig. 2.30 bottom), is now inversely proportional to a local superlattice period (in units of the survey–averaged, [001] reciprocal–lattice vector) characterizing the particular subset of superlattice repeats chosen for imaging with STM.

For purposes of precision measurement, this conceptually appealing paradigm must be further refined to take account of the asymmetry between fast– and slow–scan directions in STM, as well as the irreducible, image–to–image drift in reciprocal–lattice vector components (Fig. 2.28). We do so by re–interpreting the data in Fig. 2.31 as follows: first, assuming the superlattice wave vector remains aligned with the [001]

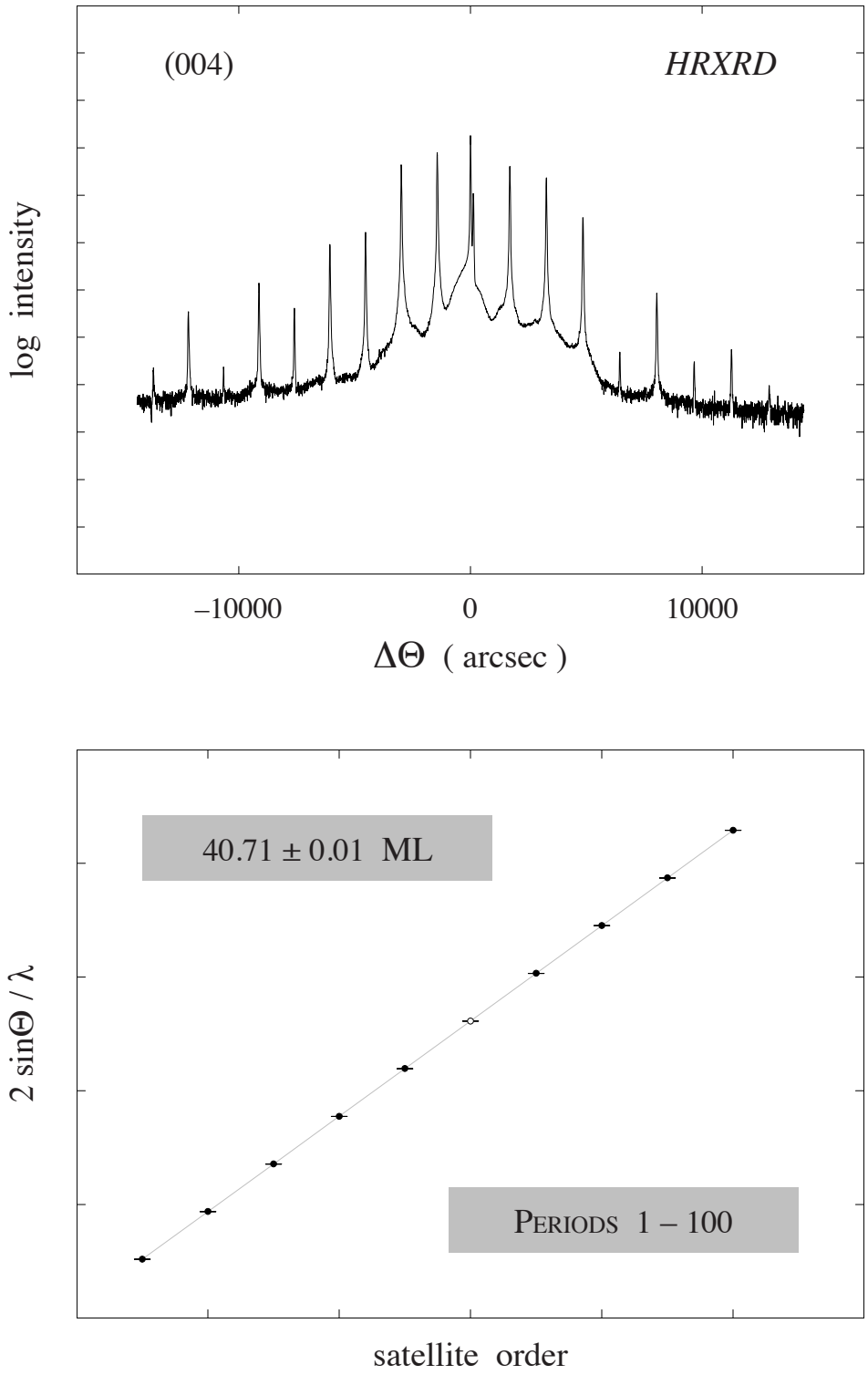


FIGURE 2.30. Conventional (004) rocking–curve (top) and corresponding Bragg plot (bottom) used to determine the SL period. The average period measured by x–ray diffraction for the entire multilayer stack is 40.71 ± 0.01 ML.

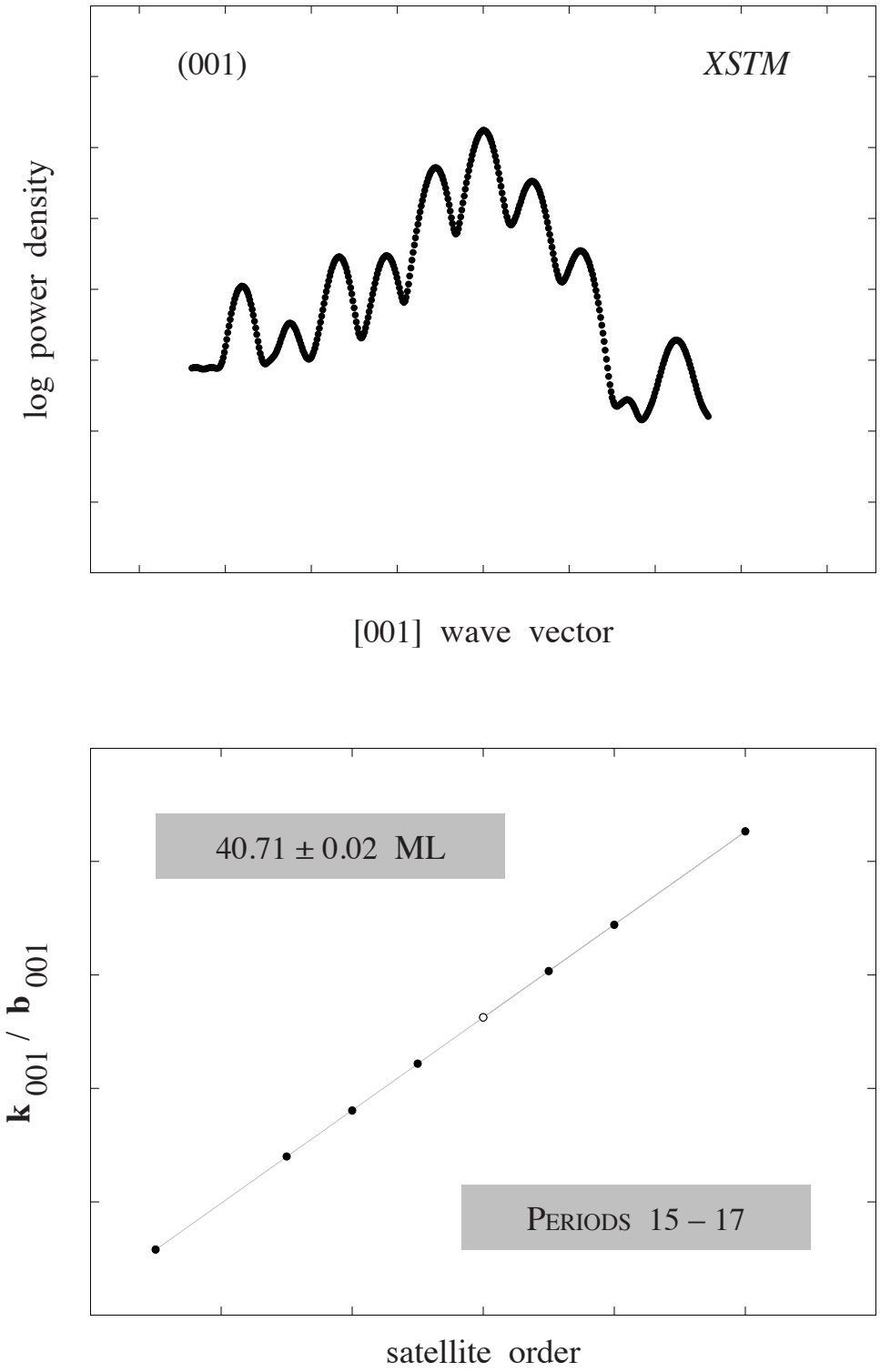


FIGURE 2.31. An [001]–section through a power spectrum (cf. Fig. 2.23) reveals satellite peaks (top) whose separation, like that of their x–ray analogs, may be used to determine the SL period (bottom). The local period for 3 repeats measured with STM is 40.71 ± 0.02 ML.

growth direction (i.e., the superlattice does not tilt considerably) there is a one–to–one correspondence between our plots in Fig. 2.31 and similar, independent plots of the *x*– and *y*–components for the same data; whereas either component will provide reasonable average values and reasonable error bounds, for the resulting local superlattice period, it should come as no surprise the *x*–, or fast–, scan data are especially stable and reproducible. Next, to eliminate any effect of drift in the reciprocal–lattice vector components themselves, the Bragg analysis in Fig. 2.31 (bottom) is applied on an image–by–image (rather than survey) basis and the resulting image–by–image periods subsequently averaged; this procedure underlies the striking similarity of *x*– and y–component results (typically better than 1% in any given image) in spite of the notably different drift characteristics (Fig. 2.28) of these two measurements.

The local period obtained from a handful of repeats thus analyzed with STM (40.71 ± 0.02 ML, Fig. 2.31 bottom) is in stunning agreement with the HRXRD value (40.71 ± 0.01 ML, Fig. 2.30 bottom) deduced for the entire multilayer stack. Additional support for the accuracy and reproducibility of our STM period measurement comes from a parallel analysis of the vertical survey in Fig. 2.27; the 100–repeat result, here, is again 40.71 ± 0.02 ML[4].

[4] This vertical survey is also amenable to a direct count of every deposited monolayer throughout the 100 repeats; the tallied superlattice period of 40.66 ± 0.04 ML is in very good agreement with our DFT method. The uncertainty in this measurement corresponds to the ambiguity of correctly detecting the first and last surface monolayer with STM. Since only every second bulk (001) plane projects dangling bonds into the vacuum following cleavage for zincblende {110} facets, the uncertainty is doubled to 4 bulk monolayers out of 100 repeats.

The pair of lateral surveys assembled in Fig. 2.32 provides further benchmarks for the internal consistency of our period measurements. These surveys were acquired over the same subset of repeats, and nominally identical surface areas, of the same cleave on consecutive days, albeit with slightly different tunnel voltages; a shared surface imperfection, indicated in Fig. 2.32, precisely anchors the spatial registry of the two mosaics. The image–by–image periods separately deduced from the last ten forward and reverse scans of each survey are plotted side–by–side in Fig. 2.33. The error bars for each measurement correspond to the uncertainty in the slope from a least–squares line to the Bragg–like plot (see Fig. 2.31, bottom) assuming the uncertainty in each satellite peak is given by the sum of squared residuals per degree of freedom [51].

We draw two, empirically–important conclusions from Fig. 2.33. First, within each data set, periods independently derived from forward and reverse scans are remarkably close (with a relative difference of ~ 0.5%), as judged either by their respective image–by–image Bragg errors or against the values themselves; that observation suggests these errors may legitimately be reduced by averaging forward and reverse measurements. Second, the point–by–point scatter in Fig. 2.33 is strongly correlated between surveys; that observation, in addition to demonstrating the extraordinary instrumental stability associated with our experimental protocols, points to the sample itself, rather than measurement uncertainty, as its source.

This insight is underscored in Fig. 2.34, where the image–by–image, forward–plus reverse–scan period averages are plotted versus in–plane displacement, and one sees the two, independent STM surveys track one another nearly perfectly. The

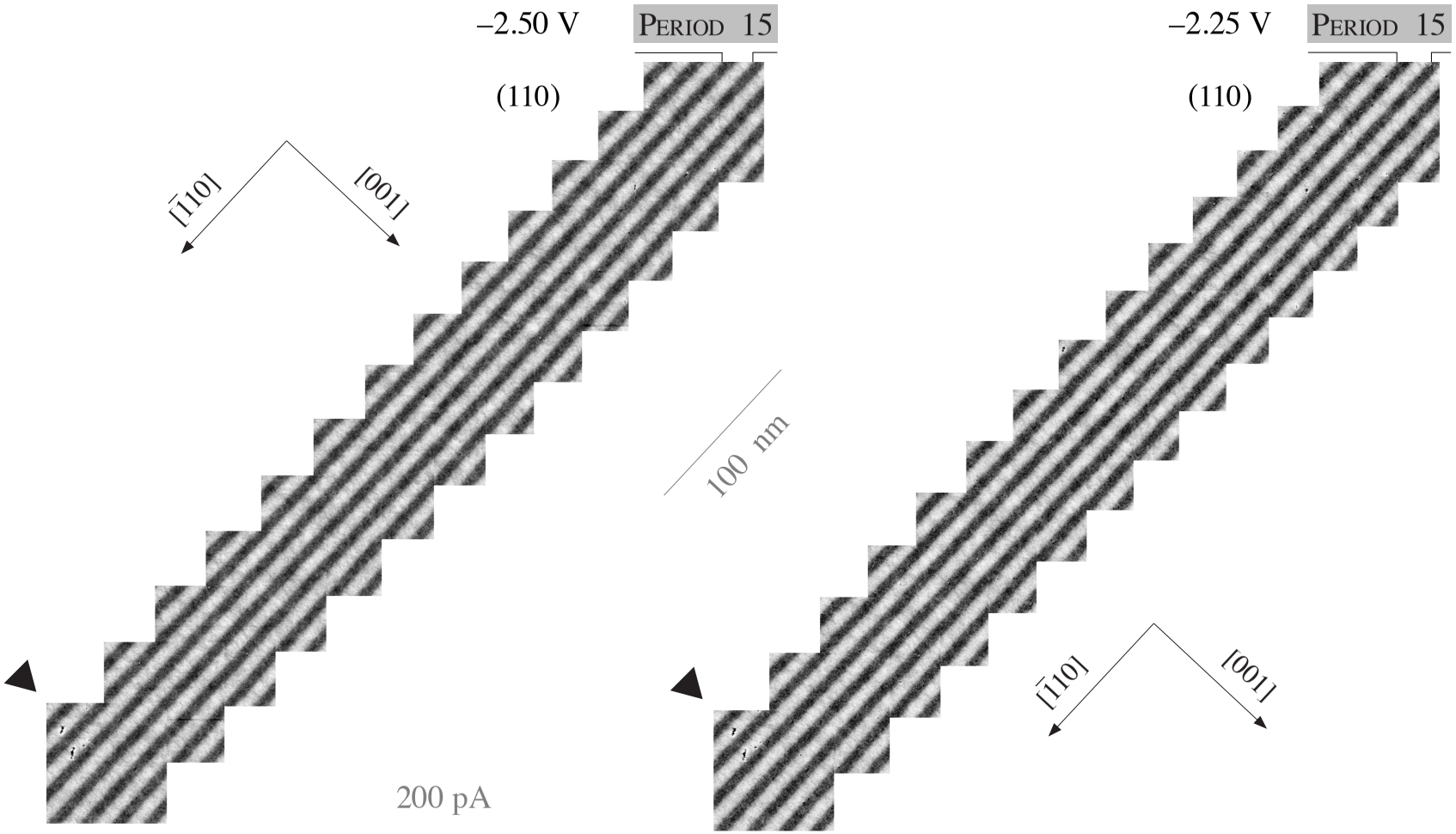


FIGURE 2.32. Device–scale, atomic–resolution surveys acquired over the "same" area of a SB InAlAs / InGaAs superlattice on consecutive days at two distinct tunnel voltages –2.50 V (left) and –2.25 V (right). The caret indicates a cleavage–induced imperfection demonstrating that nominally the "same" area was imaged in both surveys. Growth direction is from top–left to bottom–right in both surveys.

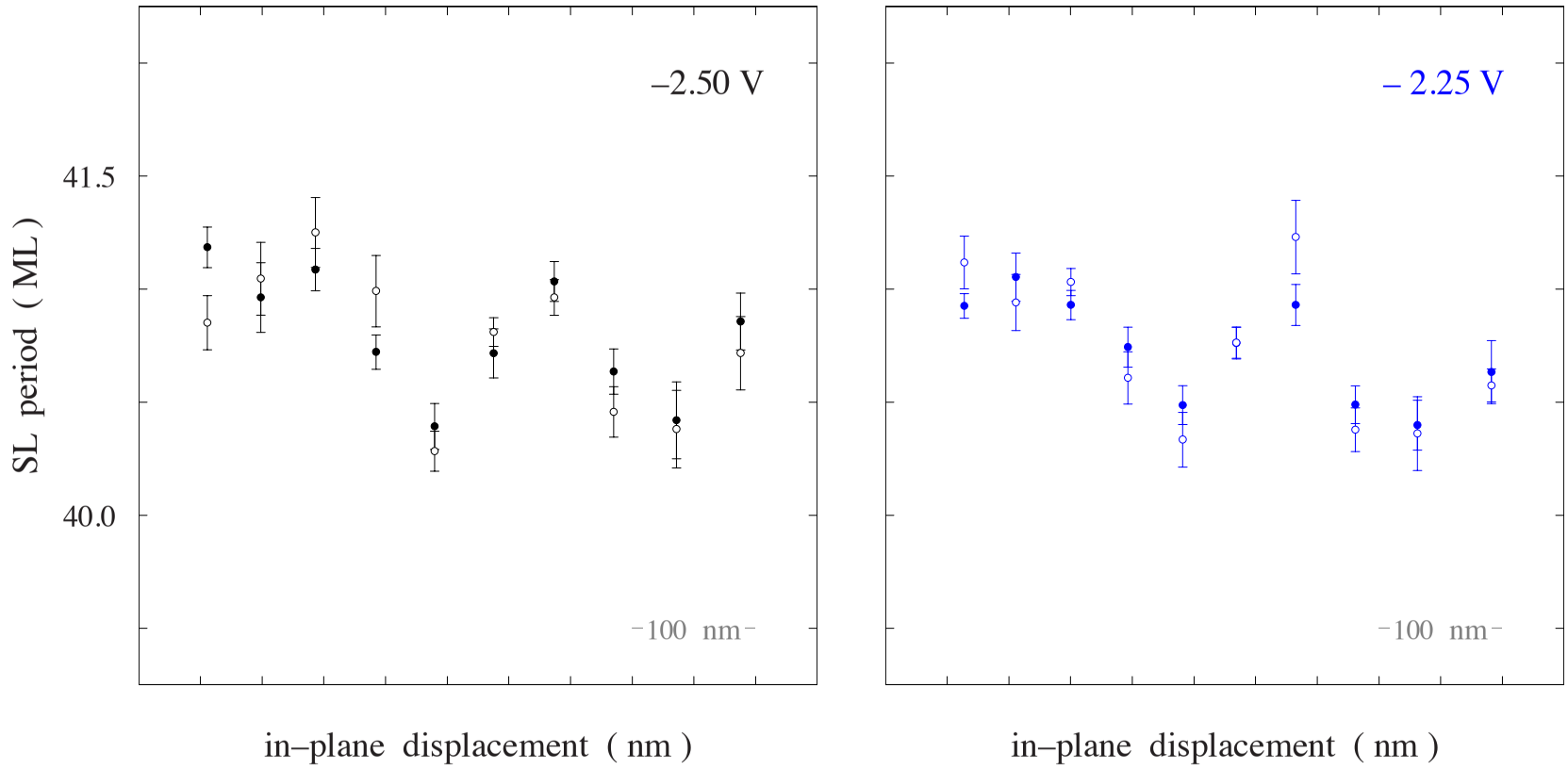


FIGURE 2.33. Image–by–image SL period determined with STM for the last 10 images from the lateral surveys in Fig. 2.32. Results from forward– (closed circles) and reverse– (open circles) scans agree very well suggesting the two may be averaged on an image–by–image basis. Vertical axis has a suppressed zero to emphasize, both, the agreement between forward– and reverse–measurements as well as the similarity between the two surveys in terms of an overall shape.

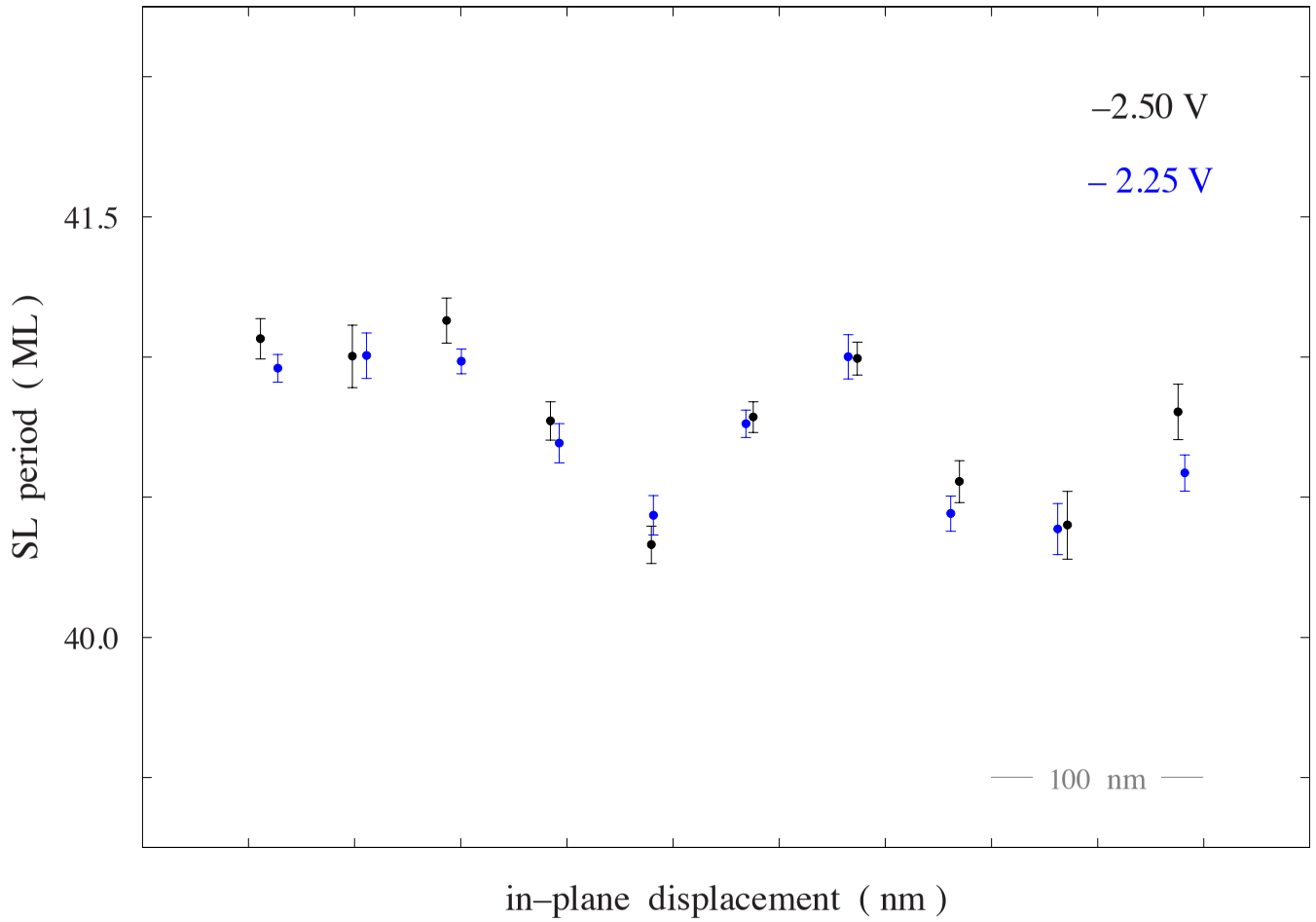


FIGURE 2.34. Image–by–image SL period averaged from forward– plus reverse–scans. The strong similarity of the two plots convincingly demonstrates that our STM–deduced SL period measurement is indeed reproducible and that the same area was in fact probed twice in consecutive surveys validating the high–level of command developed in our STM navigation protocol. Vertical axis has a suppressed zero to emphasize the agreement in measurements between the two independent surveys.

reduction in Bragg error achieved by averaging likewise makes it clear these otherwise reproducible, point–by–point fluctuations lie well outside any reasonable bounds predicted on the basis of our random measurement uncertainties alone.

Finally, because it serves as a preview of closely related questions we will be forced to confront in Chapter IV, a word of caution is in order concerning the quoted uncertainties for our survey–averaged, local period measurements. These are calculated by (uncorrelated) propagation of the individual image–by–image uncertainties (i.e., the Bragg uncertainties). This approach assumes these uncertainties are known, correctly calculated, and that the underlying process is random so they may be treated as statistically independent quantities. A more general method for characterizing the uncertainty in a sample mean value is to compute the standard error in the mean, which follows from the ratio of the spread (standard deviation) in the sampling distribution, itself, to the square root of the number of measurements [51]. This approach sidesteps assumptions concerning statistical independence and remains valid for correlated data (i.e., samples drawn from an underlying physical process whose statistical properties are correlated). The two approaches are equivalent when the physical samples involved are indeed random and independent, but the standard error of the mean provides a more representative measure if they are, in fact, correlated.

Translating this discussion into the language of STM, although the Bragg error provides an excellent estimate of the period precision attainable in any one image (per the agreement between surveys demonstrated in Fig. 2.34), its propagation is not (necessarily) representative of the actual spread in the sampling distribution, which

might well depend on other physics. The uncertainty in local superlattice period computed from a given number of STM images is then better portrayed by the standard error in the mean. For data such as that considered here – where all of the Bragg errors are essentially equal – the relative magnitude of these two estimates of uncertainty will be in direct proportion to the ratio of the rms sample spread to single–image Bragg error.

Summary

We have demonstrated here that it is possible to transform a length measurement tool with well–recognized problems – STM – into an accurate one capable of reliably determining the local superlattice periodicity with a method analogous to x–ray diffraction. This advance was achieved by the development of standardized navigation routes that considerably reduce piezo creep and systematically offer images with minimum distortion. Despite piezo non–idealities asymmetrically affecting fast– and slow– scan directions, both *x*– and *y*– components of the reciprocal–lattice vectors provide reasonable period measurements with a relative difference of ~ 1%. This observation is clearly important since any spatial investigation regarding interface fluctuations and lateral correlations necessarily involves both components.

These developments relied on the monumental efforts necessary to substantially improve vacuum levels, which in turn produced a pristine habitat where freshly exposed Al–rich surfaces remained clean for several days permitting extended periods of STM exploration. Equally essential was the overall improvement and understanding of crucial cleaving parameters necessary to reliably produce atomically–flat cleaves on InP–based

semiconductors that facilitated the exploration in cross section of a more challenging strained–balance heterostructure.

All of these technical improvements were necessary to accurately measure length scales essential to any description of interface roughness. We will now turn our attention on how to clearly identify and isolate heterojunctions from STM images; and then, in Chapter IV, we will address the subtleties associated in the statistical description of interface variance roughness for stochastic processes in the presence of lateral correlations.

CHAPTER III

INTERFACE PROFILES

Introduction

In this chapter we carefully describe and critically examine the sophisticated image–processing techniques needed to experimentally identify, extract, and digitize individual heterojunction profiles from cross–sectional STM images of a strain–balanced InAlAs / InGaAs superlattice. In succeeding chapters we examine the statistical properties of these experimentally–defined heterojunctions, first, from the perspective of their roughness amplitudes (or, more properly, mean segment variances) and, second, from the perspective of their spatial–frequency power spectra, whose Fourier transform yields the corresponding autocorrelation function.

Our discussion follows three main lines: first, we develop a systematic approach based on the strong layer contrast exhibited by a strain–balanced superlattice to identify, delineate, and extract experimental heterojunction profiles from the cross–sectional STM images. This approach relies on a well–established statistical algorithm to identify a single, threshold grey–level in the contrast histogram that most effectively separates quantum wells from quantum barriers; the heterojunction profiles determined this way offer a logical definition for the confinement potential's spatial boundaries, consolidating all contributions influencing the band–offset discontinuities between wells and barriers into a single, experimentally–accessible property throughout the cleavage–exposed cross

section. To the extent experimentally feasible, we next examine the role of key physical parameters, such as tunneling voltage, on the heterojunctions so obtained. Finally, constraints imposed by the requirement of equidistant–sampling to combine the power spectra for appropriately–defined ensembles of individual heterojunctions force a careful re–examination, and balanced assessment, of inescapable geometrical non–idealities, (already brought to light in Chapter II) whose end product is minimally–distorted, interface profiles (in units of pixels, or nominal Å) suitable for subsequent analysis.

Heterojunction Identification

The statistical analyses presented later draw on three, device–scale, atomic–resolution surveys obtained from independent cleaves of nearby dies (Fig. 2.3). Two of these surveys specifically targeted the same subset of early superlattice periods in complementary (110) (Fig. 3.1, left) and (1–10) (Fig. 3.1, right) cross sections, thereby eliminating temporal evolution (but not wafer location) as a potentially significant variable. Cursory examination of late superlattice periods, though not pursued here, suggests such temporal evolution may indeed be an important factor; our early–period comparison of the as–grown structure in orthogonal cross sections, on the other hand, affords a much–needed perspective on the potential role of growth–plane anisotropy.

The exceptional quality of the survey mosaics assembled in Fig. 3.1 may be better appreciated by referring to Fig. 3.2, where individual STM images are reproduced. As we will see in the section immediately following, this incredible resolution – essential for providing reliable, atomic–length–scale standards by way of the in–plane

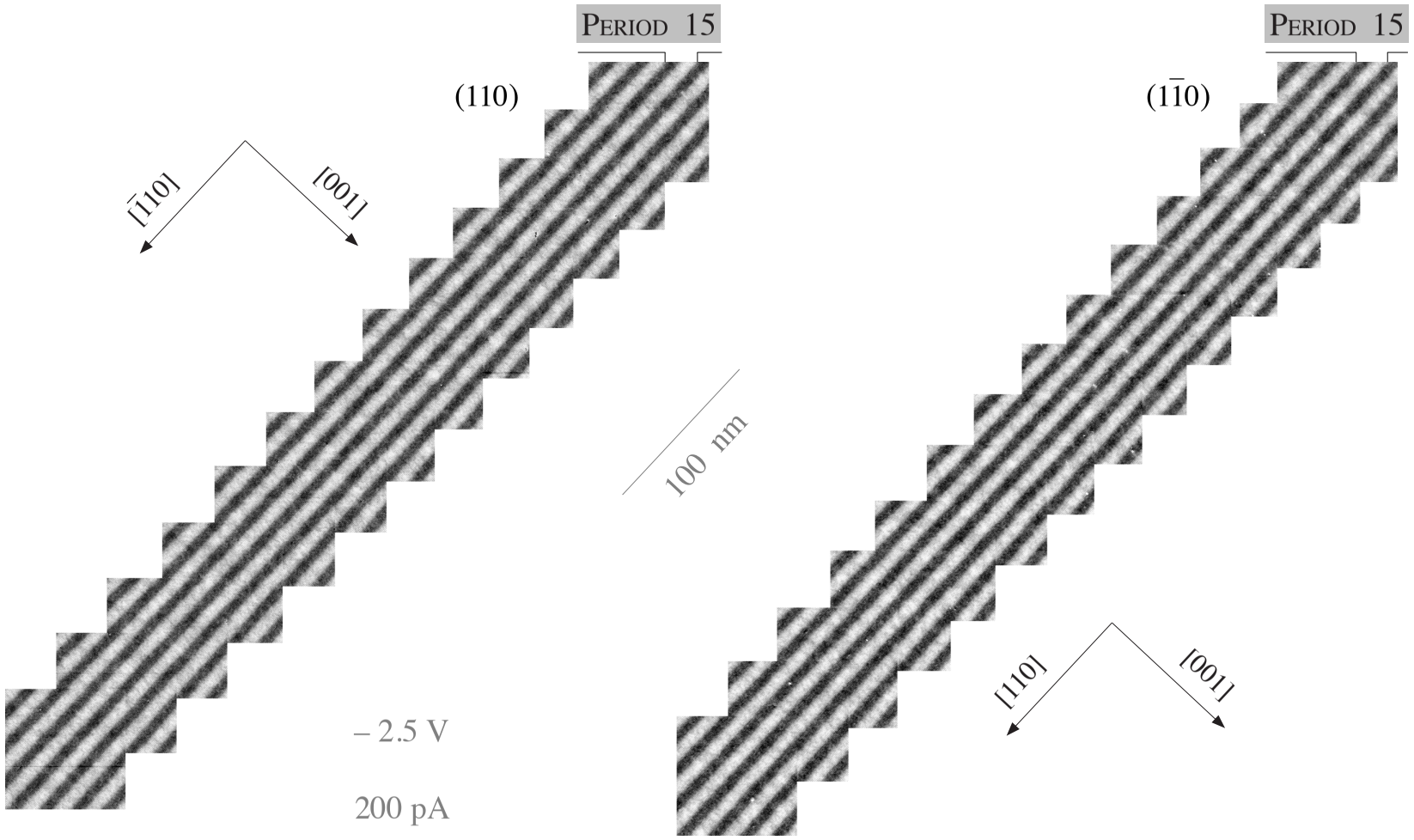


FIGURE 3.1. Device–scale, atomic–resolution surveys acquired over identical periods of a strain–balanced InAlAs / InGaAs superlattice in orthogonal (110) and (1–10) cross sections. Quantum well and barrier components of the superlattice are identified in Fig. 3.2. Growth direction is from top left to bottom right in each case.

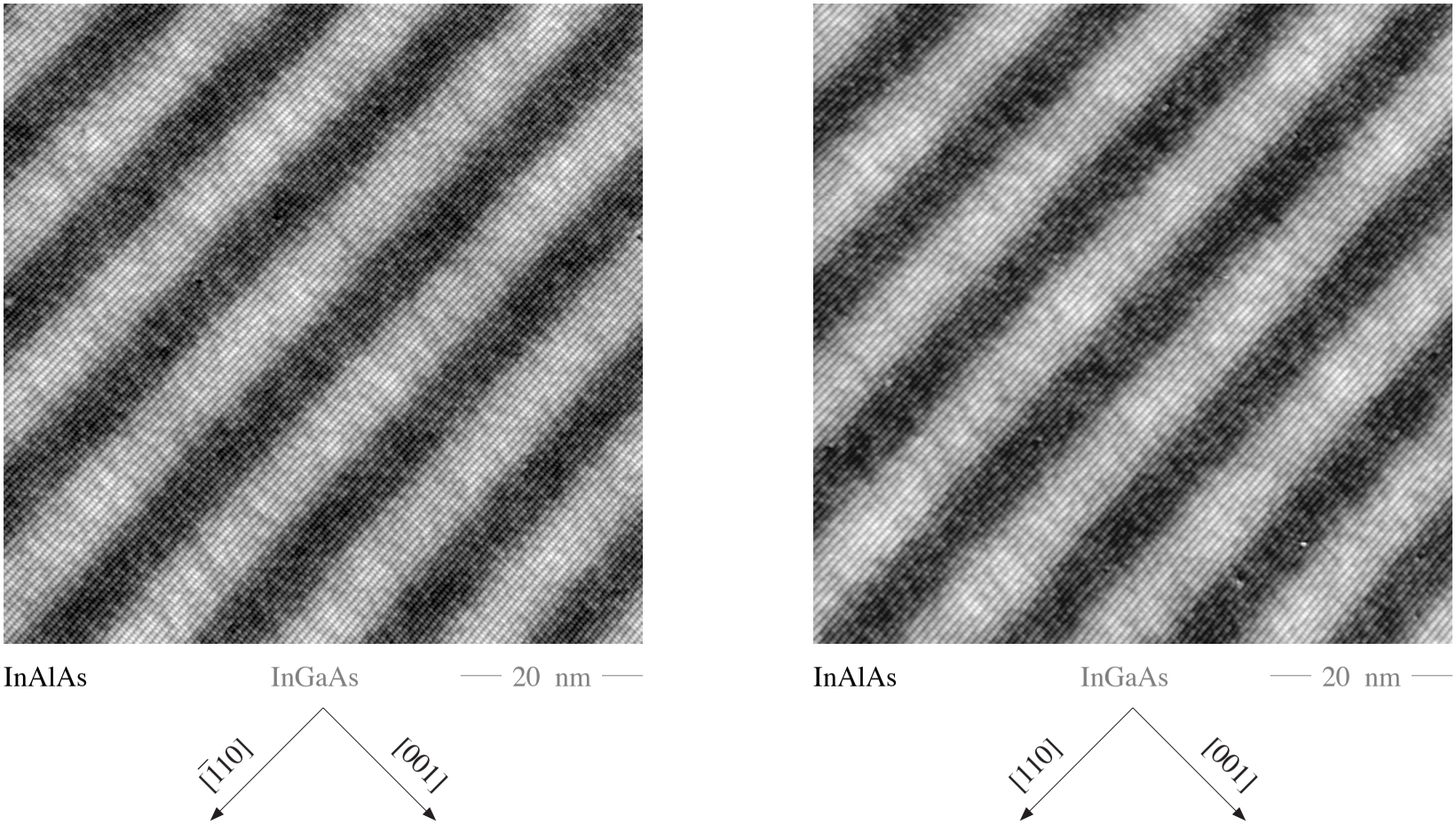


FIGURE 3.2. Atomic–resolution, anion–sublattice images from the (110) (left) and (1–10) (right) surveys in Fig. 3.1. Growth direction is from top left to bottom right in each case.

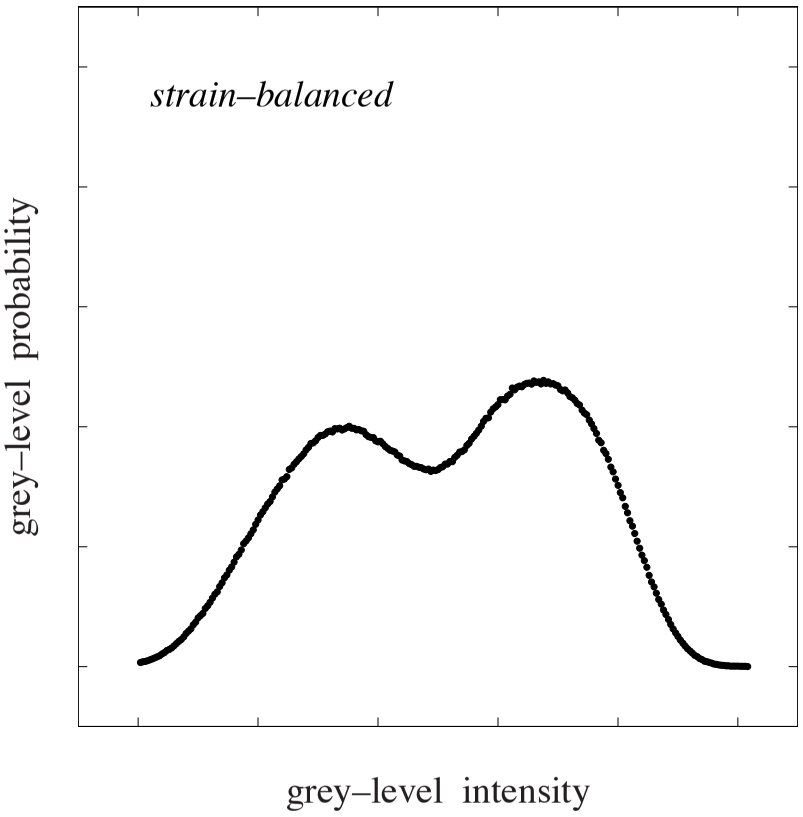
reciprocal–lattice vectors – is something of a double–edged sword from the all–important standpoint of spatially delineating individual quantum wells and barriers.

In this section we describe the image–processing algorithms used to define and, later, extract the interfaces between adjoining quantum wells and barriers by digitally isolating the two, constituent layers from one another. As already mentioned, the strain–balanced superlattice (Fig. 3.3, right) exhibits a well–defined, bimodal, grey–level histogram (Fig. 3.3, left), suggesting that a clean, statistical partition between quantum wells and quantum barriers might be feasible.

The idea, simply, is that there is a grey–level value – in the vicinity of the local minimum situated between the two peaks in Fig. 3.3 (left) – that "best" separates the two sub–populations. A well–established statistical algorithm, developed by Otsu [52], precisely addresses this question, independent of any assumptions concerning the sub–population statistics (Gaussian or otherwise). The algorithm is easily implemented, and widely employed[1], to determine the grey–level threshold (Fig. 3.4, left) that achieves "optimal" spatial segmentation (Fig. 3.4, right) for a given contrast histogram.

As indicated in Fig. 3.4, although the two–tone mask obtained by replacing all grey levels below the Otsu threshold by one color, and those above it by another (e.g., black and grey), largely succeeds in separating wells and barriers, it demonstrably fails to do so in the very regions of most interest to us – i.e., along the heterojunctions defining the transition from one material to the other.

[1] The statistical metric used here maximizes peak separability relative to peak widths, which is an analogous procedure to the Rayleigh diffraction criterion found in optics.



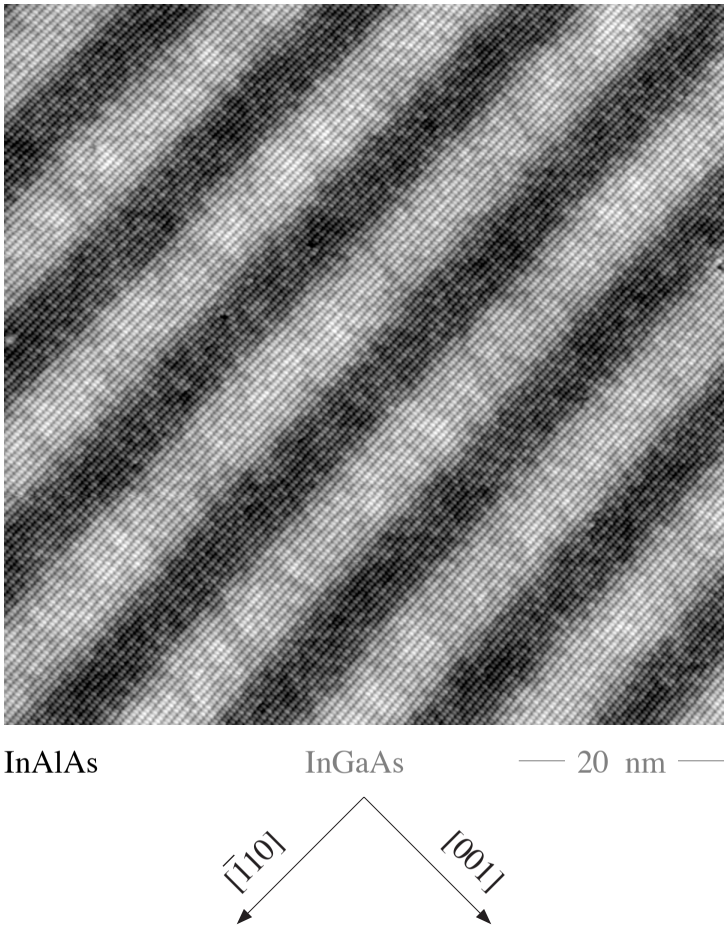


FIGURE 3.3. Representative contrast histogram (left) derived from the ensemble of atomic–resolution images (right) compromising the (110) survey illustrated in Fig. 3.1. The bimodal grey–level distribution suggests a statistical partitioning to spatially delineate well and barrier subpopulations may be possible. Growth direction is from top left to bottom right.

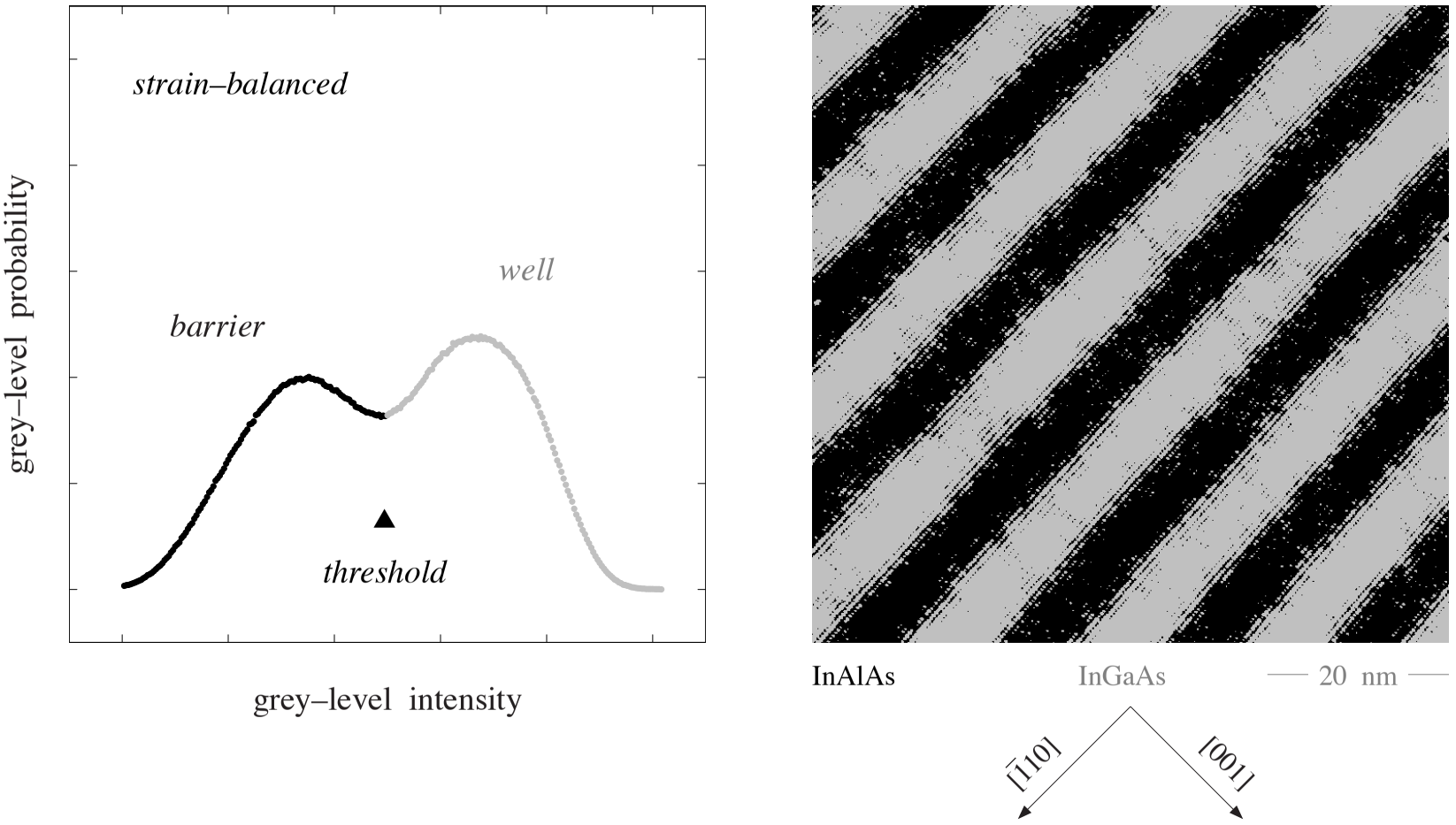


FIGURE 3.4. Contrast histogram from Fig. 3.3 (left), with the (Otsu) threshold grey level that best separates wells and barrier populations indicated. Grey–values below this threshold are set black, and those above it set grey, to create a black–and–grey mask (right) that may be used to assess the degree to which spatial segmentation has been successfully achieved. In this case, interspersed regions where the two colors mix are problematic. Growth direction is from top left to bottom right.

The reasons for this failure may be best appreciated by examining the surface sections (Fig. 3.5, bottom) from sequential quantum wells and barriers (Fig. 3.5, top). These sections expose the problem succinctly: atomic troughs, belonging to InGaAs layers on one side of the interface (dashed line), lie below the Otsu threshold and therefore set to black, whereas atomic crests, belonging to InAlAs layers on the other side of the interface, lie above the Otsu threshold, and therefore set to grey. We thus learn a two–tone mask cannot everywhere distinguish InAlAs from InGaAs in the face of atomic corrugation amplitudes comparable to the superlattice layer contrast itself (Fig. 3.5, bottom).

Fortunately, our previous consideration of the power spectra attending such STM images (Chapter II) suggests an immediate remedy: though low– and high–frequency components – reflecting superlattice and atomic periodicities, respectively – coexist, they nevertheless populate distinct, and hence separable, regions of reciprocal space. As already noted, the power spectrum (Fig. 3.6, top) from a strain–balanced InAlAs / InGaAs superlattice exhibits two, distinct groups of spots, one of which corresponds to the 2–D atomic–mesh reciprocal–lattice vectors (red circles) and their convolved superlattice satellites, whereas the other group, localized near the zone center, arises solely from the superlattice periodicity (blue circle). By means of a suitable “low–pass” filter in reciprocal space, one might reasonably hope to cleanly separate the two.

The specific reciprocal–space filter chosen for this purpose is a Fermi function whose Fermi "energy" lies somewhere between the origin and the [001] reciprocal–lattice vector, and whose "temperature" (in units of the Fermi energy) has been adjusted

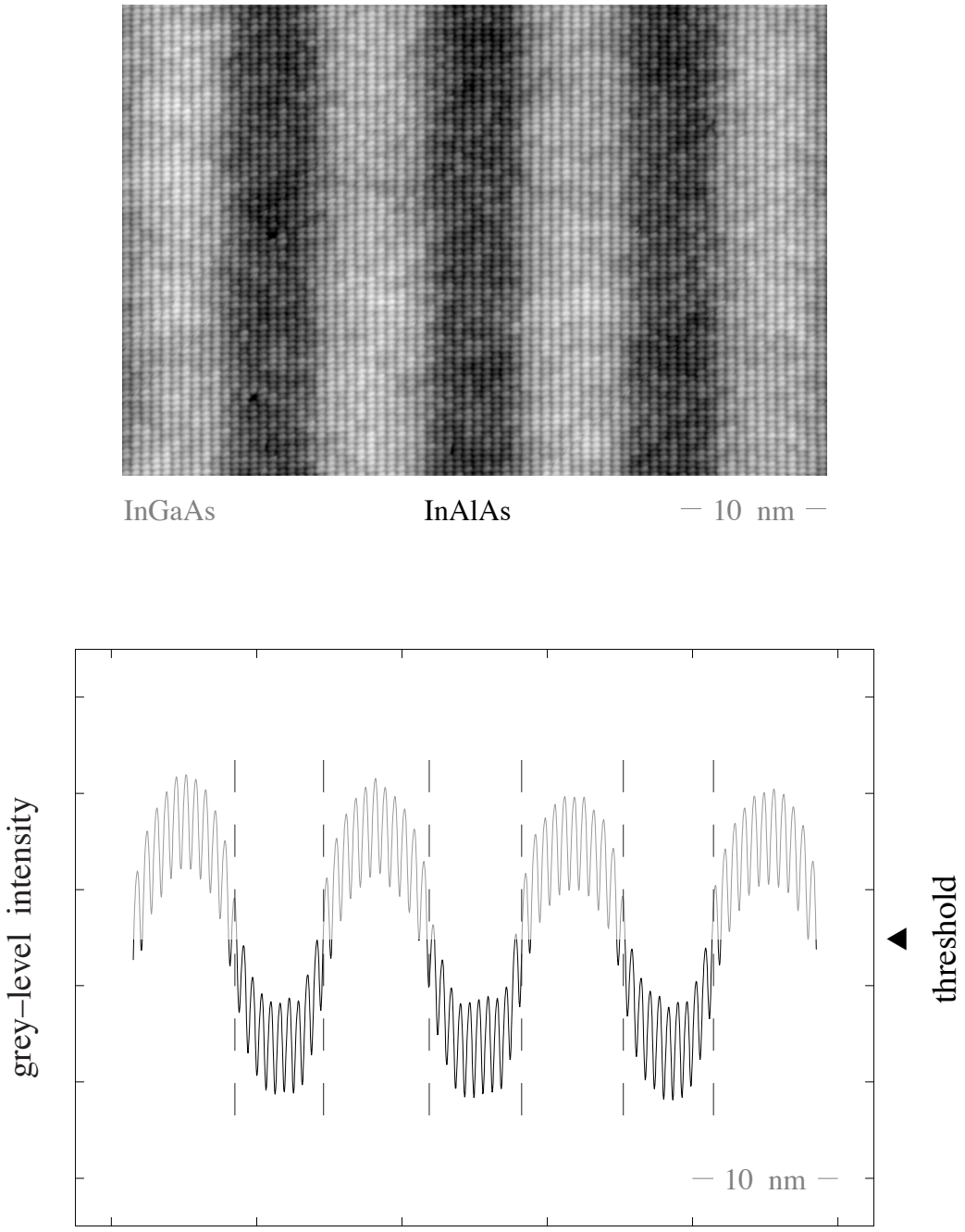


FIGURE 3.5. In–plane–average surface sections (bottom) through consecutive quantum wells and barriers (top) illustrate a “high” frequency component associated with atomic corrugation modulating the “low” frequency signal associated with superlattice layer contrast. This atomic corrugation is responsible for the two–tone intermixing at interfaces (dashed lines) that impedes successful spatial segmentation in Fig. 3.4 (right).

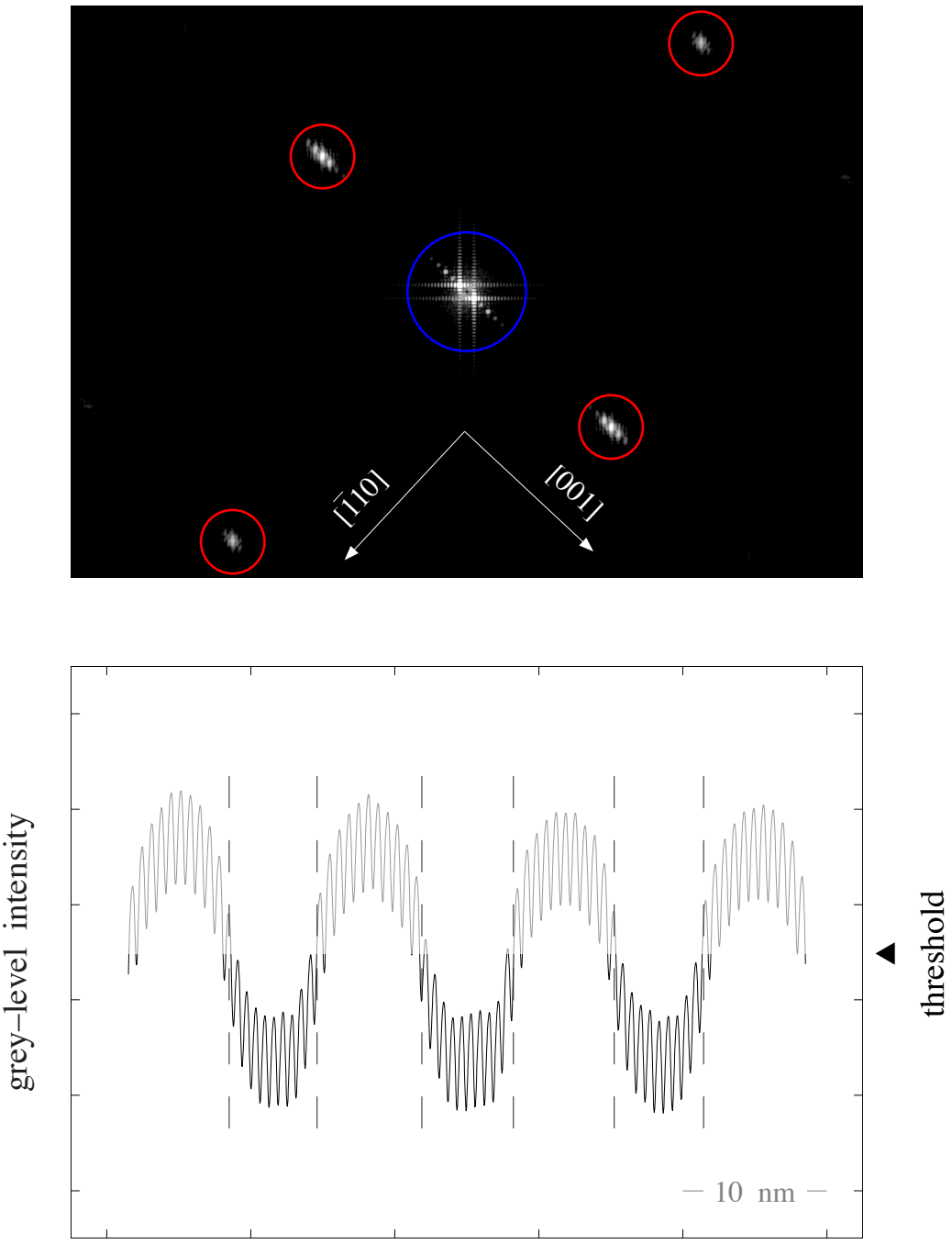


FIGURE 3.6. Ensemble–average, reciprocal–space power spectrum (top) from the atomic–resolution survey in Fig. 3.1 (left). Satellite peaks near the zone center (blue) and in the vicinity of the cleavage–plane reciprocal–lattice vectors (red) are due to the superlattice periodicity; corresponding [001] surface section (bottom), from Fig. 3.5.

to appropriately soften an otherwise abrupt "cutoff" and thereby eliminate any measurable ringing in direct space once the filtered spectrum is inverse transformed. Fig. 3.7 (top) portrays the resulting band–limited power spectrum, and Fig. 3.8 (top) the corresponding, low–pass–filtered, STM image. The resulting surface sections (Figs. 3.7 and 3.8, bottom) are, essentially, corrugation free, so only the superlattice contrast modulation remains, and it is at once clear (via the light and dark highlighting superposed on our filtered section) that an Otsu threshold may now be employed to cleanly partition quantum wells and barriers.

Filtered and unfiltered contrast histograms are compared in Fig. 3.9. Both are bimodal, but it is obvious that filtering (right) results in better–defined sub–populations, with improved discrimination between the two, assumed classes. This improvement is statistically validated by the sharpness of the so–called separability curve (grey), which is constructed while exhaustively searching for the grey–level value that maximizes the statistical discrimination between populations in Otsu's algorithm. The grey level at maximum separability is the threshold value later adopted for the binary masks then used to spatially discriminate between quantum wells and quantum barriers.

The digital transformation from grey–scale filtered STM images to binary masks provides a one–of–a–kind empirical tool for mapping out the heterojunctions effectively delineating each constituent layer. Representative masks from orthogonal cross sections

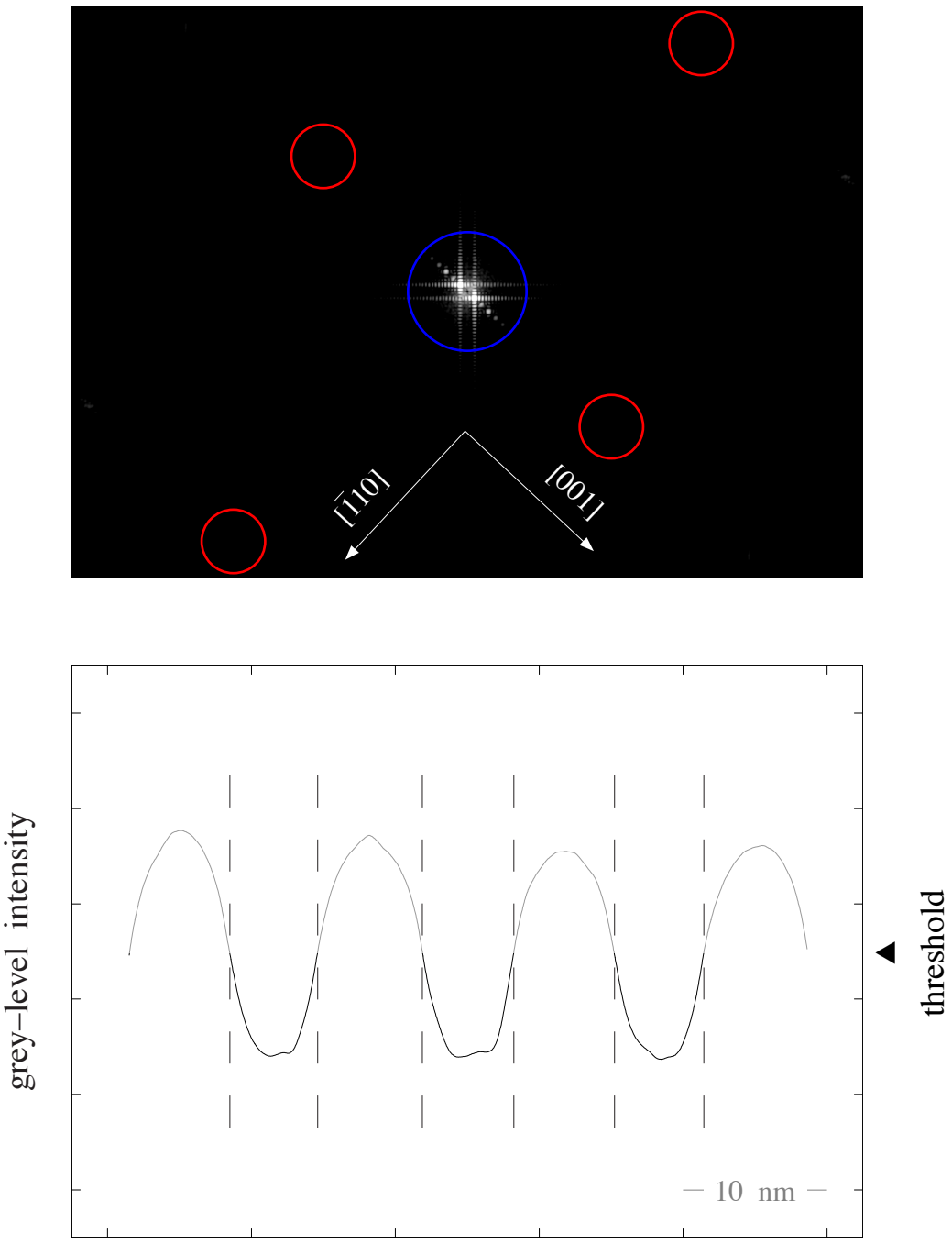


FIGURE 3.7. Bandwidth–limited, reciprocal–space power spectrum after multiplication with a low–pass, “Fermi” filter (top) and corresponding [001] surface section following inverse transformation back into direct space (bottom). Atomic corrugation has been effectively removed from the cleavage–surface profile and the two–tone intermixing previously seen in Fig. 3.5 (bottom) eliminated as a consequence.

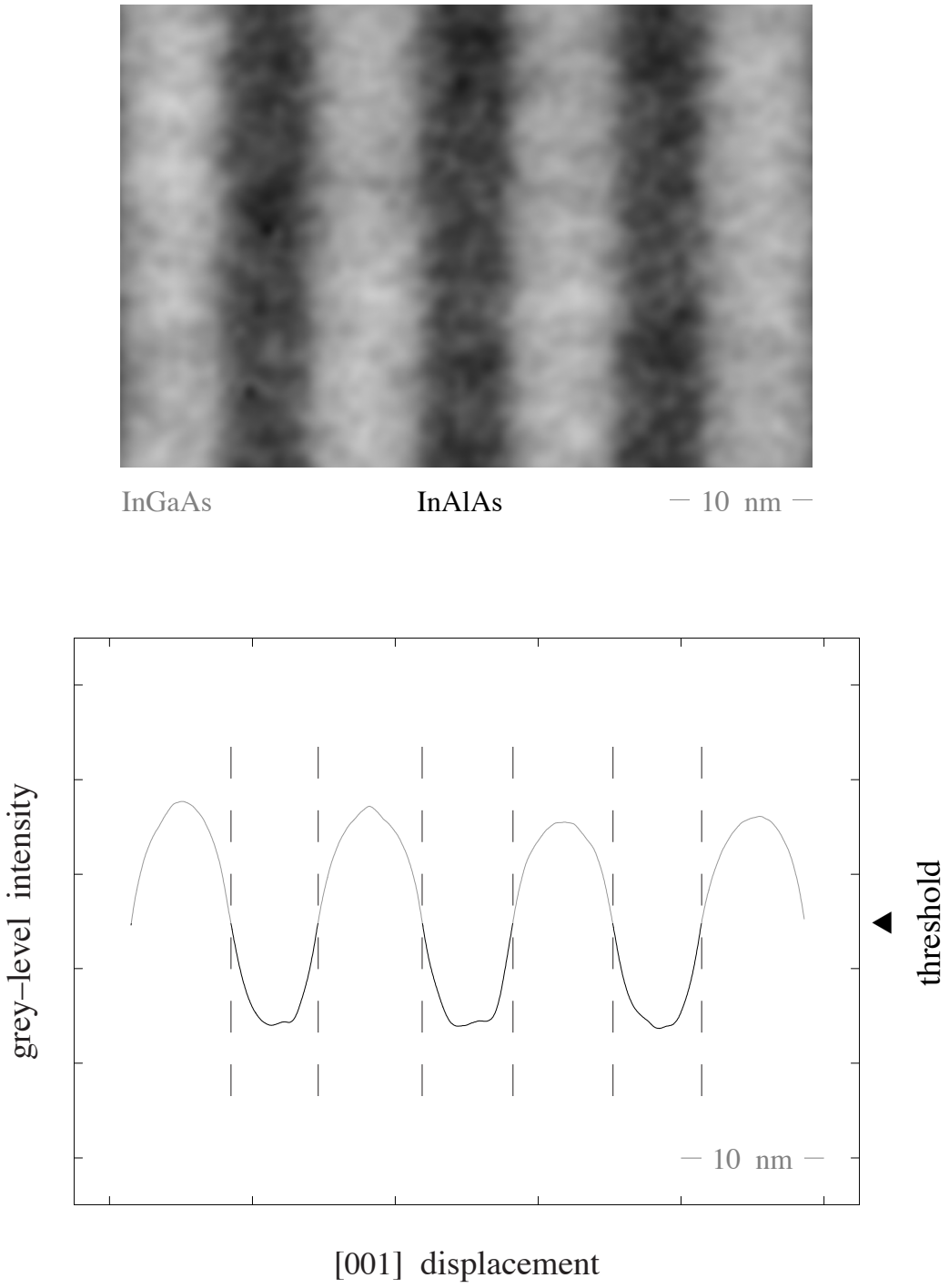


FIGURE 3.8. Bandwidth–limited STM image obtained from inverse transformation of the low–pass filtered reciprocal–space power spectrum in Fig. 3.7 (top) and corresponding [001] surface section (bottom).

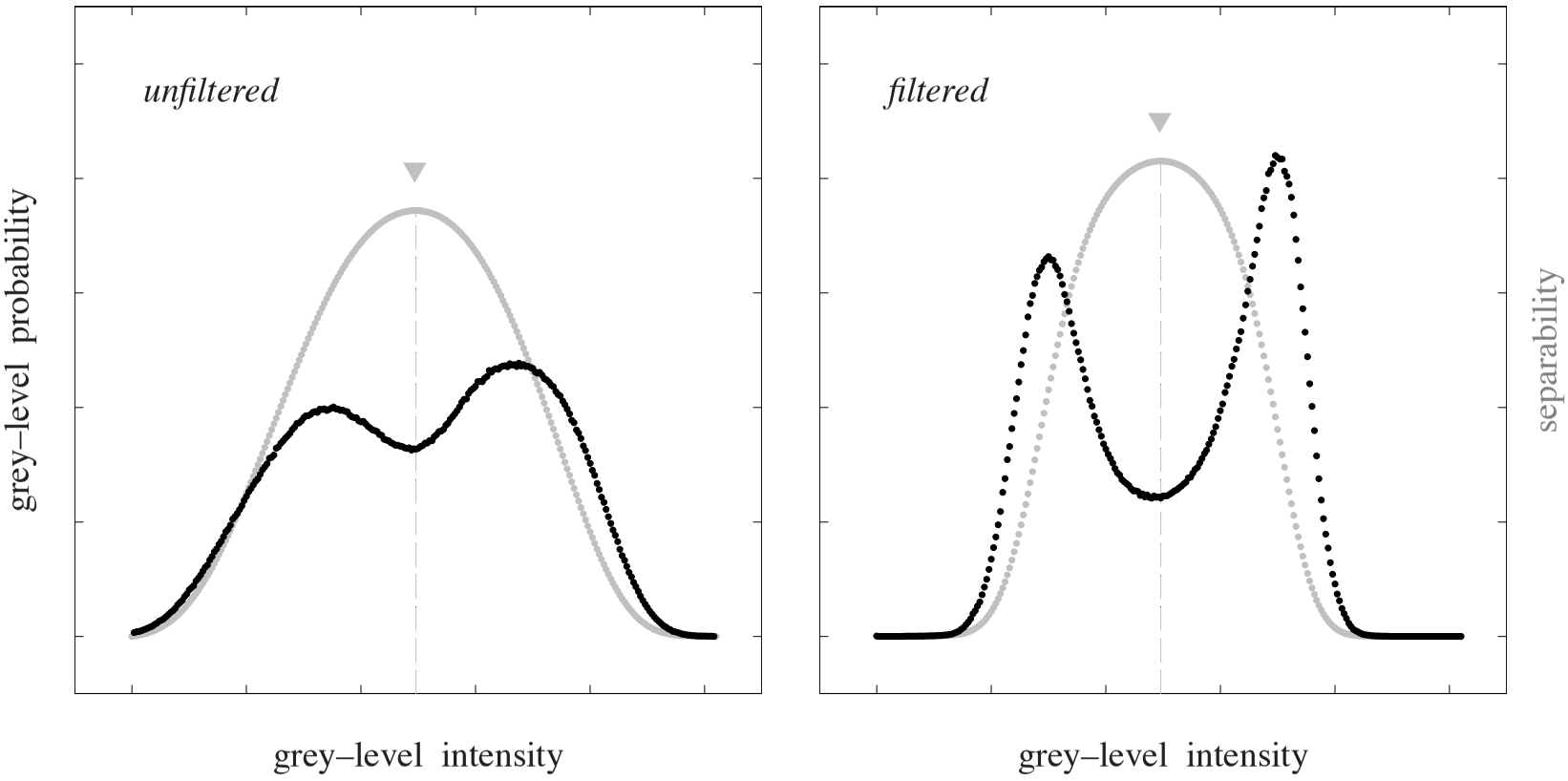


FIGURE 3.9. Ensemble–average contrast histograms assembled from unfiltered, atomic–resolution images (left), and filtered, bandwidth–limited images (right), compromising the (110) survey illustrated in Fig. 3.1. Maximum in the Otsu separability curve (grey) sets the threshold grey level (carat) subsequently adopted to delineate the interfaces in Fig. 3.10. Filtered histograms exhibit well–defined bimodal distributions, with notably sharper (more certain) separability curves than their unfiltered counterparts.

of a filtered, strain–balanced InAlAs / InGaAs superlattice are shown in Fig. 3.10 and Fig. 3.11, respectively[2].

Heterojunction Dependence on Image Processing

Before outlining how the heterojunction profiles of interest to us may be extracted from the binary masks described above, we must consider the dependence of these masks on related image processing choices, together with their possible dependence on physical conditions governing the tip–sample tunnel junction.

We begin this discussion by exploring, first, the role of the Fermi "energy"– or spatial–frequency cutoff – employed to suppress atomic corrugation in the STM images. Reduction of the Fermi filter's characterization to this single parameter pre–supposes the choice of "temperature" – or transition–width in reciprocal space – has already been made; trial and error revealed that a value comparable with the Fermi "energy" itself was entirely suitable for our purposes, so this ratio was subsequently fixed as the Fermi "energy" alone was varied to determine optimum filtering.

It was eventually discovered that, if the chosen Fermi "energy" (i.e., the blue–circle radius in Figs. 3.6 and 3.7, top) was too large, the interface contours ultimately obtained from these masks (Fig. 3.12, left) exhibited "retrograde" behavior: regions where these contours looped back on themselves and were no longer single–valued functions of the in–plane [110] coordinate – a requirement essential for our subsequent

[2] All of the grey–level histograms and Otsu thresholds pointed to here are survey averages; however, individual images comprising the mosaic in question are masked with their corresponding threshold value.

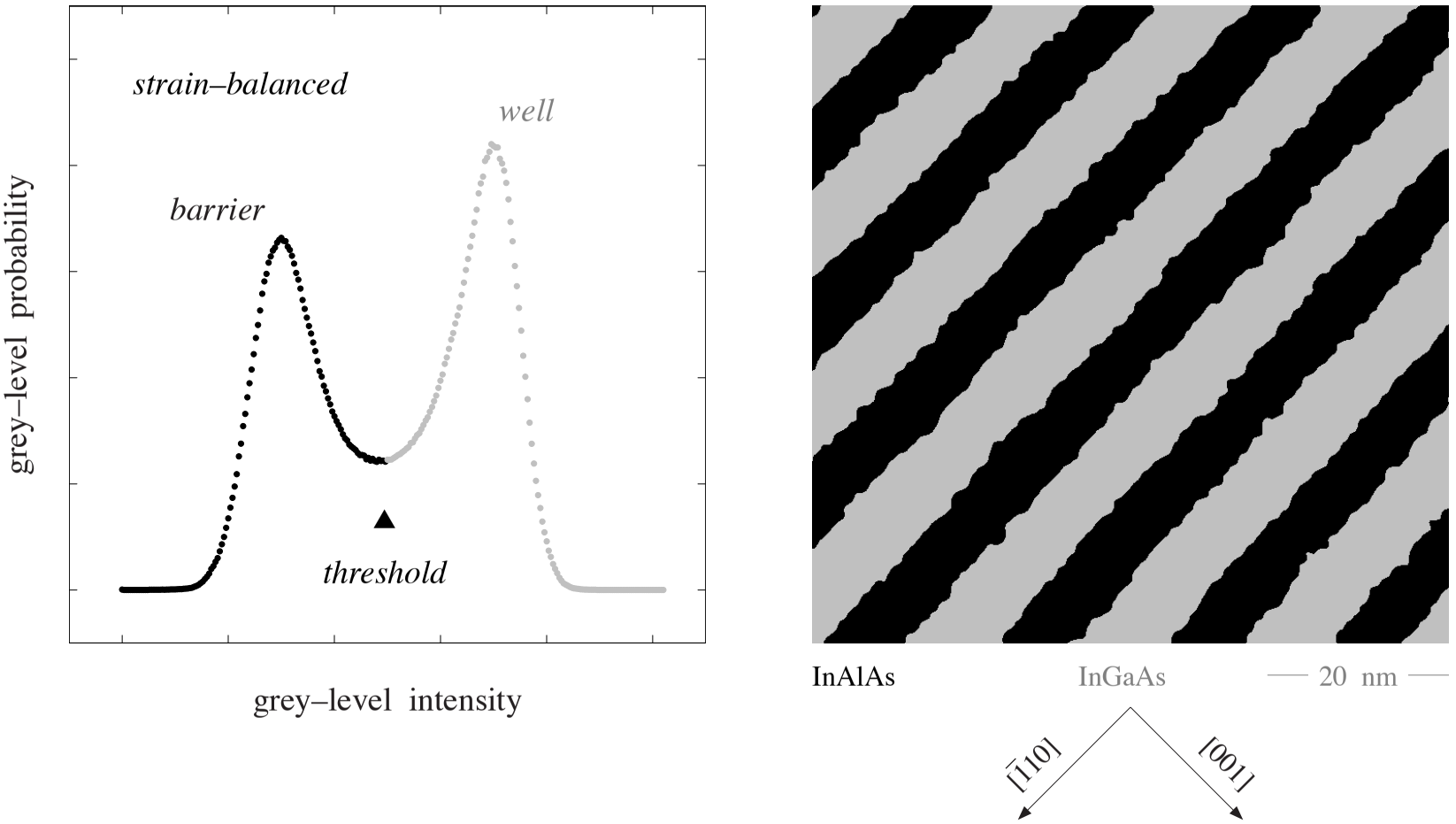


FIGURE 3.10. Contrast histogram from bandwidth–limited images of a strain–balanced InAlAs / InGaAs superlattice in (110) cross section (left) together with the Otsu threshold used to generate the binary masks (right) that empirically define the heterojunctions where adjoining wells and barriers meet. Growth direction is from top left to bottom right.

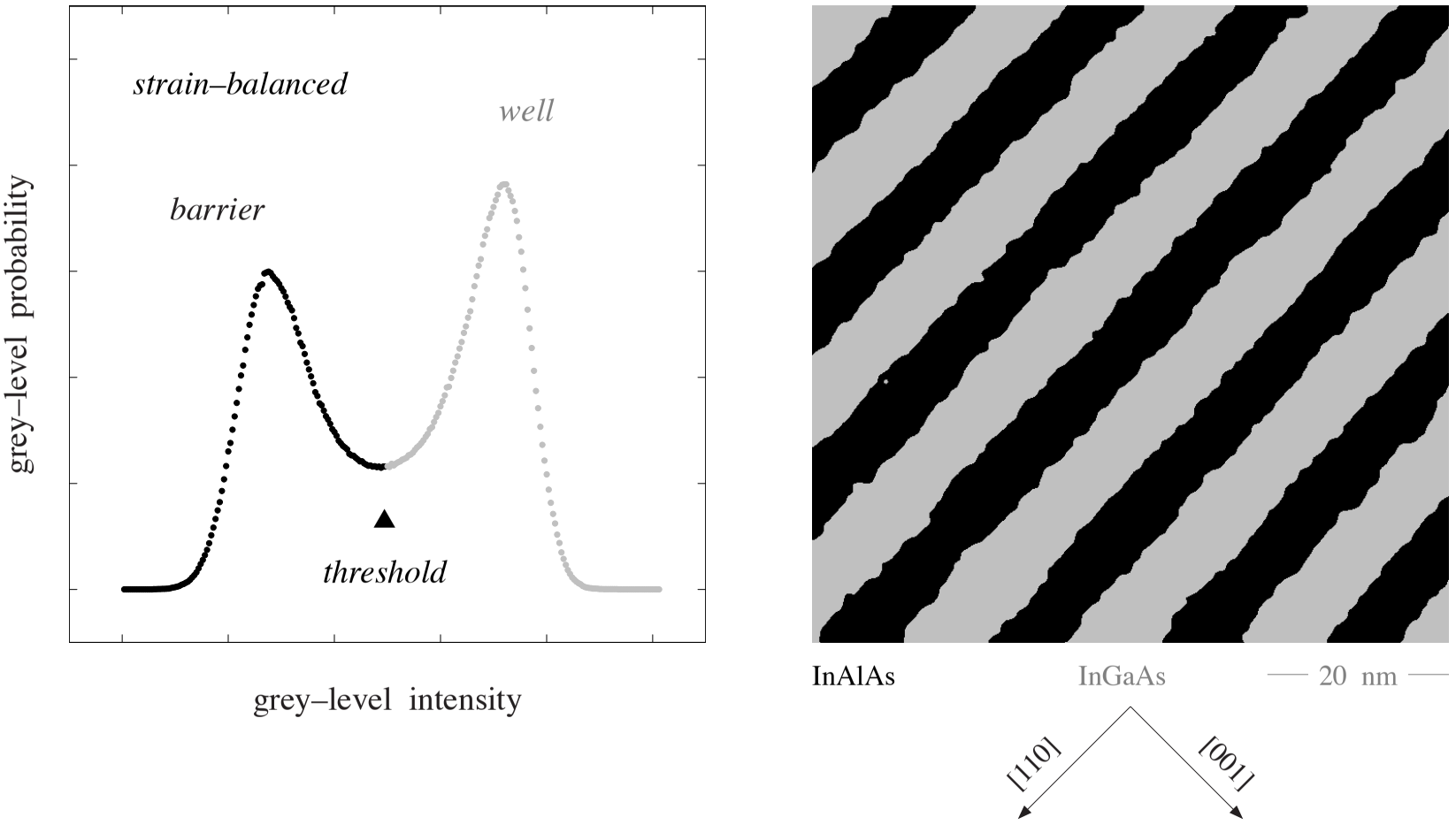


FIGURE 3.11. Same contrast histogram (left) and mask (right) as previously described in Fig. 3.10, but now in (1–10) cross section.

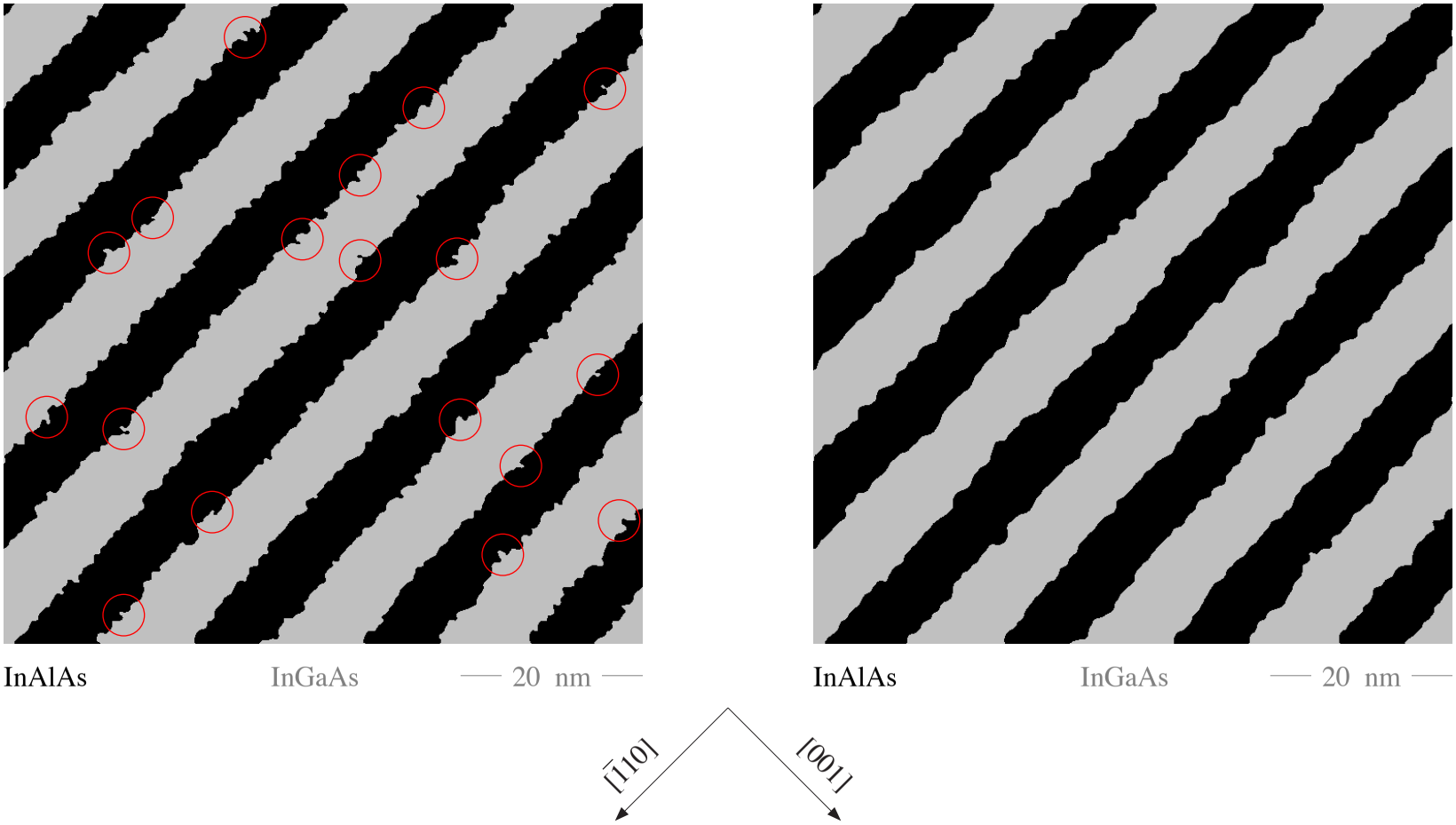


FIGURE 3.12. Influence of Fermi–filter radius (Fig. 3.7, top) on heterojunction delineation by way of binary masks. Inadequate filtering (left) leaves retrograde segments (red), where interfacial displacements in the growth direction are no longer a singled–value function of the growth–plane coordinates, behind. Stricter bandwidth limiting (right) removes this mathematically–problematic anomaly.

mathematical analysis[3]. Iterative adjustment of the filter radius in reciprocal space followed by careful visual inspection of the resulting real–space contours (Fig. 3.12, right) established our final choice for the maximum tolerable Fermi energy. The corresponding contrast histograms (Fig. 3.13), on the other hand, are (understandably) insensitive to these subtleties and thus provide no guidance concerning the optimal filter radius.

We turn, next, to the role played by the Otsu threshold differentiating "well" and "barrier" components in our optimally–filtered, grey–level histograms (Fig. 3.11). Even though the statistical reasoning behind Otsu's algorithm appears reasonable and persuasive, it is nevertheless important to examine whether the resulting binary masks (and heterojunctions) deduced from it depend strongly on this specific threshold choice. A one–grey–level uncertainty in the threshold value sets a natural scale for assessing the sensitivity of our heterojunction assignments, shifting the inferred location of any interface by at most 0.1 ML in the growth direction.

The effect of a more extreme (and perhaps unrealistic) change in threshold value of six–grey levels – corresponding to heterojunction shifts of approximately one–half monolayer in the [001] direction – was independently explored for InAlAs–on–InGaAs and InGaAs–on–InAlAs interfaces. As shown in Figs. 3.14 and 3.15, respectively, the superposed masks corresponding to Otsu threshold, and Otsu threshold ± 6 grey levels, are nearly–indistinguishable replicas of one another. Thus, while this dramatic

[3] the Fourier transform, in particular, depends on this criterion for a "well–behaved" function. On the other hand, referring back to Figs. 3.6 and 3.7, it is likewise clear that if the filter radius is naively taken to be too small, it can distort the superlattice contrast – the very object we seek to quantify – by cutting off its high frequency components.

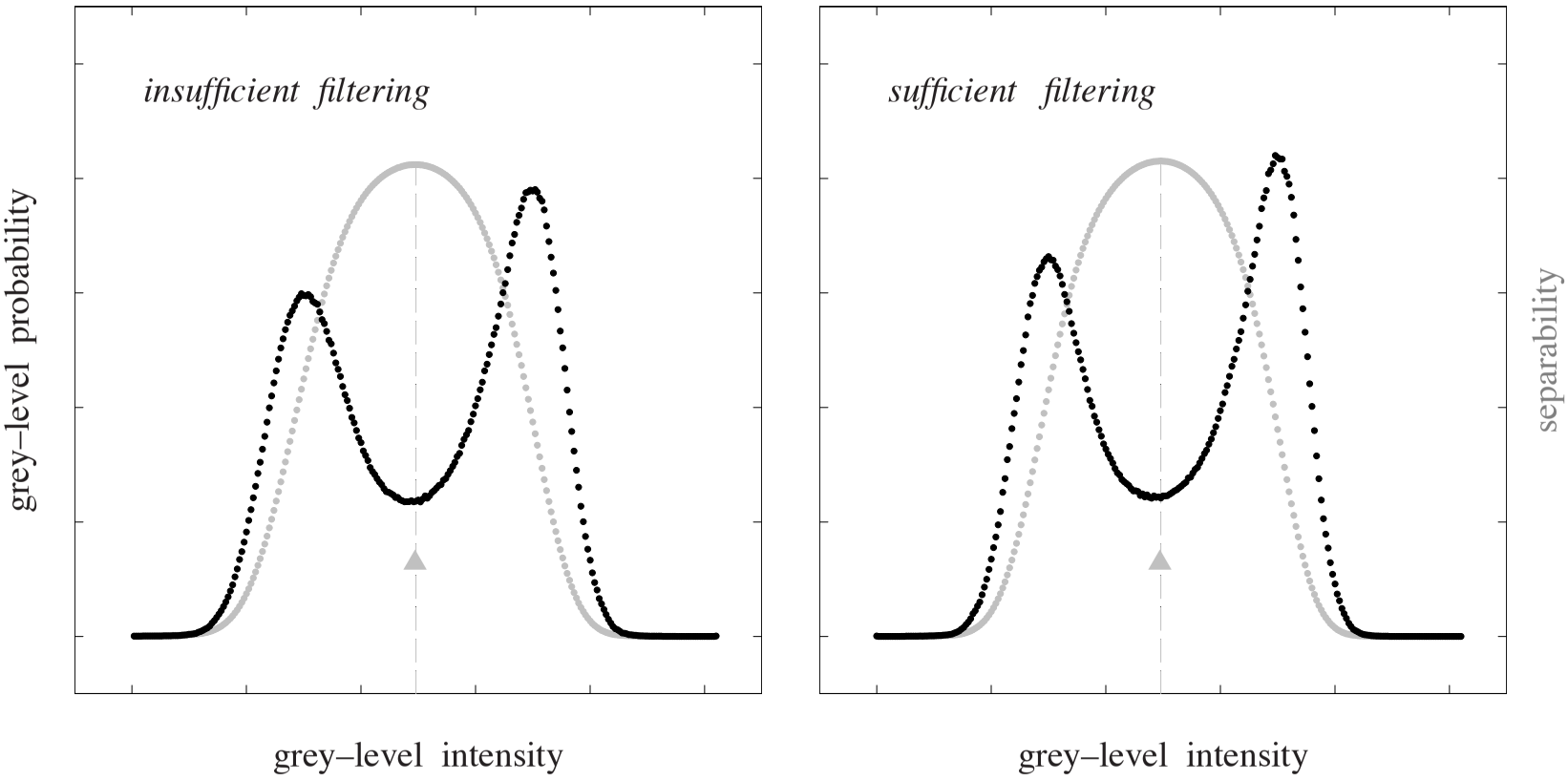


FIGURE 3.13. Ensemble–average contrast histograms used to generate Figs. 3.12 (left) and 3.12 (right), respectively. These grey–level distributions are nearly identical, offering no hint of the anomaly revealed in Fig. 3.12 by direct inspection of the masks themselves. Carats denote Otsu threshold.

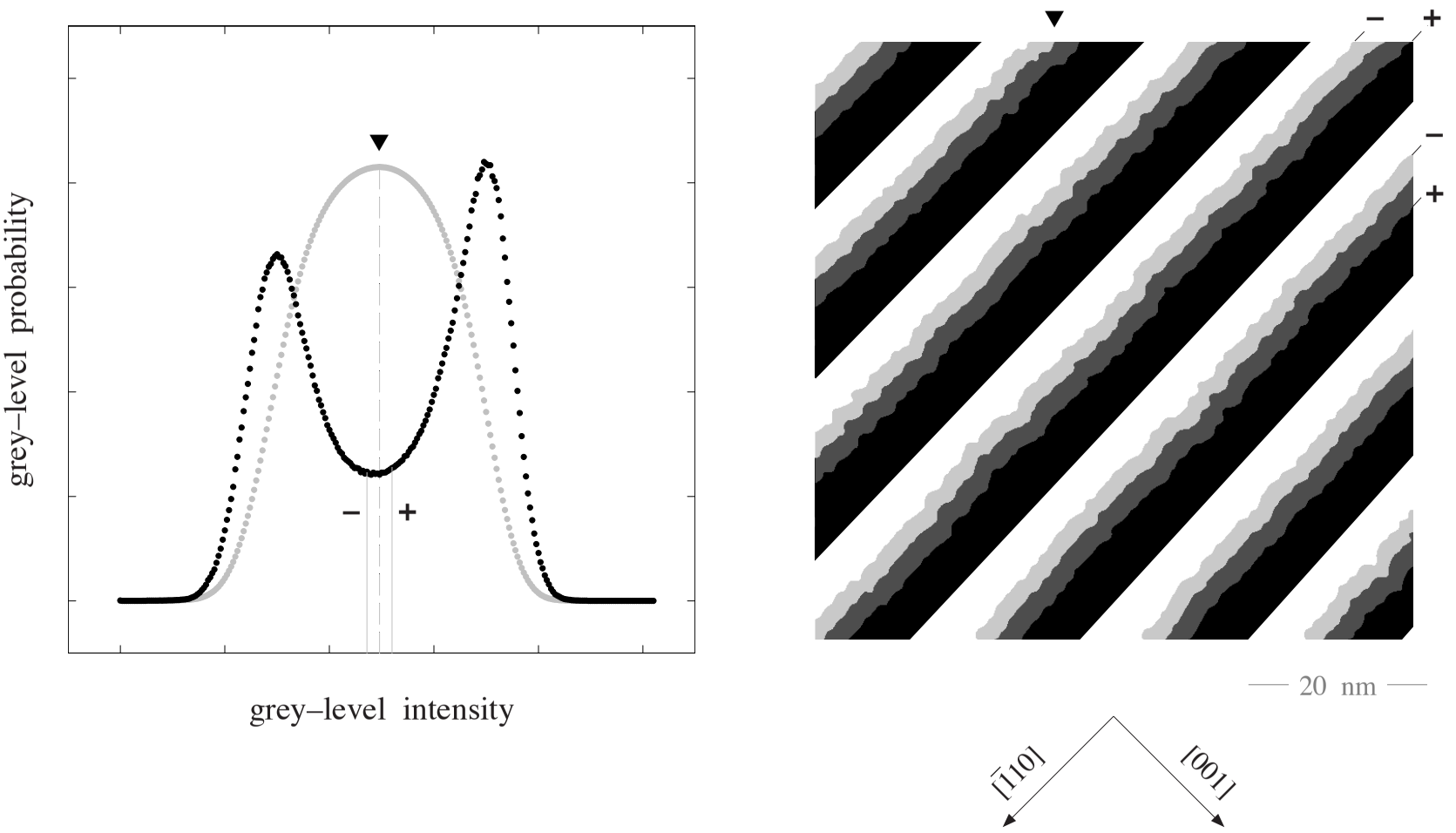


FIGURE 3.14. Threshold insensitivity of the InAlAs–on–InGaAs interface. The evident parallelism of binary masks generated from neighboring thresholds – offset, here, for clarity – demonstrates the insensitivity of our heterojunction delineation to any (reasonable) choice of threshold grey–level. Carat denotes the Otsu value, and ± denote the Otsu value ± 6 grey–levels (equivalent to ± 0.5 ML), respectively. Growth direction is top left to bottom right.

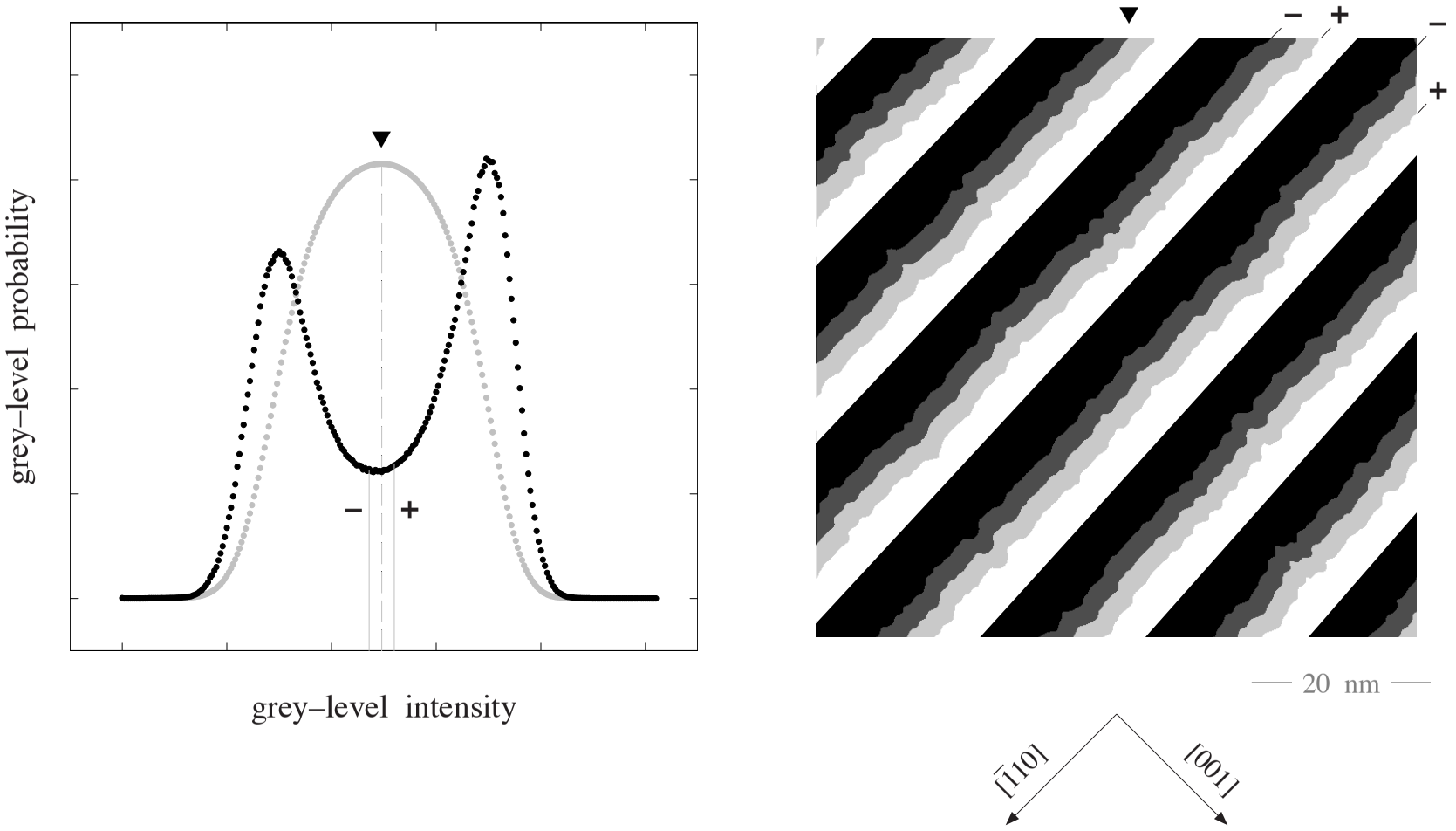


FIGURE 3.15. Threshold insensitivity of the InGaAs–on–InAlAs interface. The evident parallelism of binary masks generated from neighboring thresholds – offset, here, for clarity – demonstrates the insensitivity of our heterojunction delineation to any (reasonable) choice of threshold grey–level. Carat denotes the Otsu value, and ± denote the Otsu value ± 6 grey–levels (equivalent to ± 0.5 ML), respectively. Growth direction is top left to bottom right.

modification understandably alters the superlattice–period fraction assigned to wells and barriers respectively (the well–to–barrier ratio), it introduces no significant change to the interface profiles themselves.

We turn attention, finally, to the grey–level intensity assigned to each pixel of an STM image. This pixel intensity reflects point–to–point changes in vertical height as the STM tip is rastered over the surface and its *z*–piezo displacement monitored with a 16–bit analog–to–digital converter (ADC). Much of that dynamic range is taken up with a small, residual tilt between the cleaved surface and *x*–, *y*–piezo plane nominally perpendicular to the tip. This misorientation is further compounded by the typical non–idealities associated with piezo–ceramic scanning mechanisms already discussed at length in Chapter II, including piezo creep, which may introduce spurious surface curvature.

These concerns regarding the STM data are, in principle, addressed by digitally subtracting a suitable background from each image. A linear, least–squares fit will take care of the residual slope from any misalignment with respect to the cleavage plane, whereas quadratic, or higher–order, fits may be employed to remove overall bowing, or more complex varieties of curvature respectively, from the image. The appropriateness of various least–squares, polynomial–background subtraction algorithms up to fourth order was systematically investigated through their effect on the associated binary masks, which turn out to be more sensitive to these factors than the images themselves.

That sensitivity is especially acute around the edges of a binary mask, and comes to notice once consecutive masks from a given survey are overlaid. Planar (left) and

quadratic (right) background subtractions applied to the same pair of masks are compared in Fig. 3.16; a red circle focuses attention on a typical region influenced by this difference in polynomial order. A similar comparison, involving the same pair of masks but now including cubic (left) as well as quartic (right) terms in the fit, is shown in Fig. 3.17. It is clear from the sequence of polynomial–background–subtractions considered in these two figures that a quadratic fit offers the best transition between (overlaid) consecutive masks, whereas higher, or lower, orders introduce unphysical discontinuities at, or near, mask corners.

One last point concerning our grey–level assignments remains. Calibration of the *z*–piezo against known, mono–atomic step transitions sets a nominal sensitivity of 0.02 Å per bit for the ADC employed to monitor tip–height adjustments. Since the mean vertical *z*–separation (after background–subtraction) between InAlAs and InGaAs layers in the SB superlattice studied here is of the order of 1 Å, or only 50 ADC LSBs (Least Significant Bits), most of the 8–bit (or 256–level) color range conventionally adopted for commercially–available graphics software [53] is underutilized. To take better advantage of the available dynamic range, our originally–integer STM data is converted to floating point, and, following background subtraction, multiplied by a constant scale factor (two, in all cases considered here) before re–conversion to integer; the end result is that the dispersion of native grey values about their mean scales proportionately, thereby improving the layer contrast and histogram separability.

Side–by–side comparison of native (left) and grey–scale amplified (right) STM images is presented in Fig. 3.18, with the corresponding, Fermi–filtered contrast

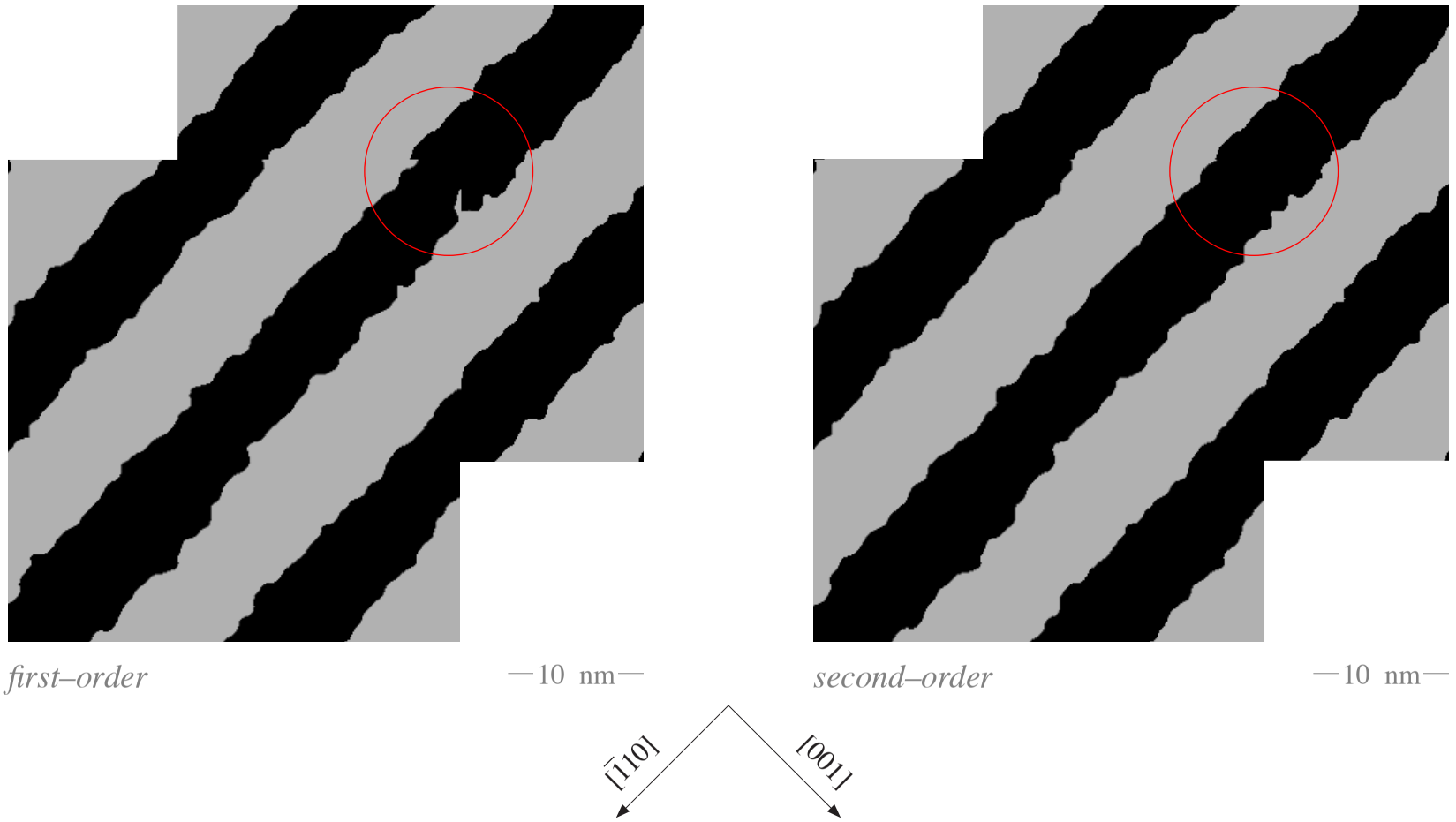


FIGURE 3.16. Effect of (low–order) polynomial–surface–background subtraction on heterojunction delineation at overlapping masks. First–order (planar) background subtraction introduces strong edge discontinuities (encircled in red), but second–order (quadratic) subtraction (right) does not, when independently applied to the same images. Growth direction is from top left to bottom right.

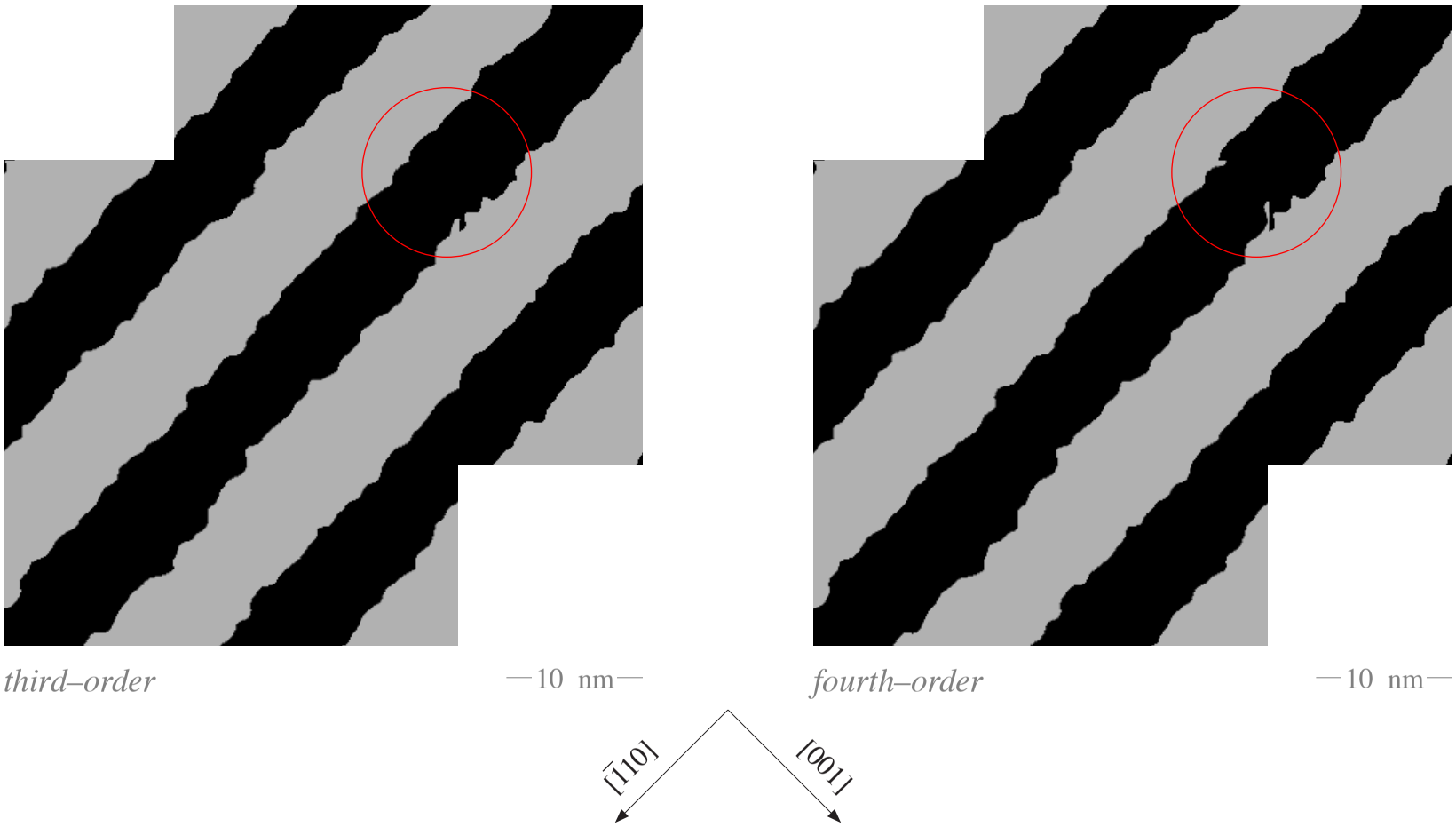


FIGURE 3.17. Effect of (high–order) polynomial–surface–background subtraction on heterojunction delineation at overlapping masks. Third– (cubic) or fourth–order (quartic) background subtractions independently applied to the same images produce undesirable edge discontinuities (encircled in red) similar to those seen with linear order alone (Fig. 3.16 left). Growth direction is from top left to bottom right.

histograms compared in Fig. 3.19. Every one of the InAlAs / InGaAs superlattice images and histograms presented in this thesis has been similarly rescaled.

Heterojunction Dependence on Physical Conditions

Two, experimentally–accessible image acquisition parameters control the purely electronic interaction between an STM tip and the local state density over a cleavage–exposed surface: tunnel current and tunnel voltage. In practice, the combined demands of uninterrupted atomic resolution and tip stability over the ten(s) (of) hours or so required to complete one (or more) micron–long STM survey(s) severely restrict the accessible parameter space; these limitations are often further exacerbated by the unique requirements posed by two, electronically–distinct materials in a periodically–modulated structure such as a superlattice. Thus, a typical, tip–sample tunnel junction exhibits "sweet spots" that are not nearly as "elastic" as one might hope. One familiar, and particularly frustrating, example is the overwhelming stability advantage universally enjoyed by the STM tip when imaging with negative (filled–state) versus positive (empty–state) sample bias.

A fixed tunnel current of 200 pA was adopted for all of our experiments. This value represented a pragmatic compromise between the typically lower currents (~ 100 pA) preferred to avoid any possible tip–induced damage to III–V materials, and higher ones (~ 300 pA) that (for the same voltage) accentuate atomic corrugation as well as superlattice contrast. Although the range of experimentally–accessible tunnel voltages

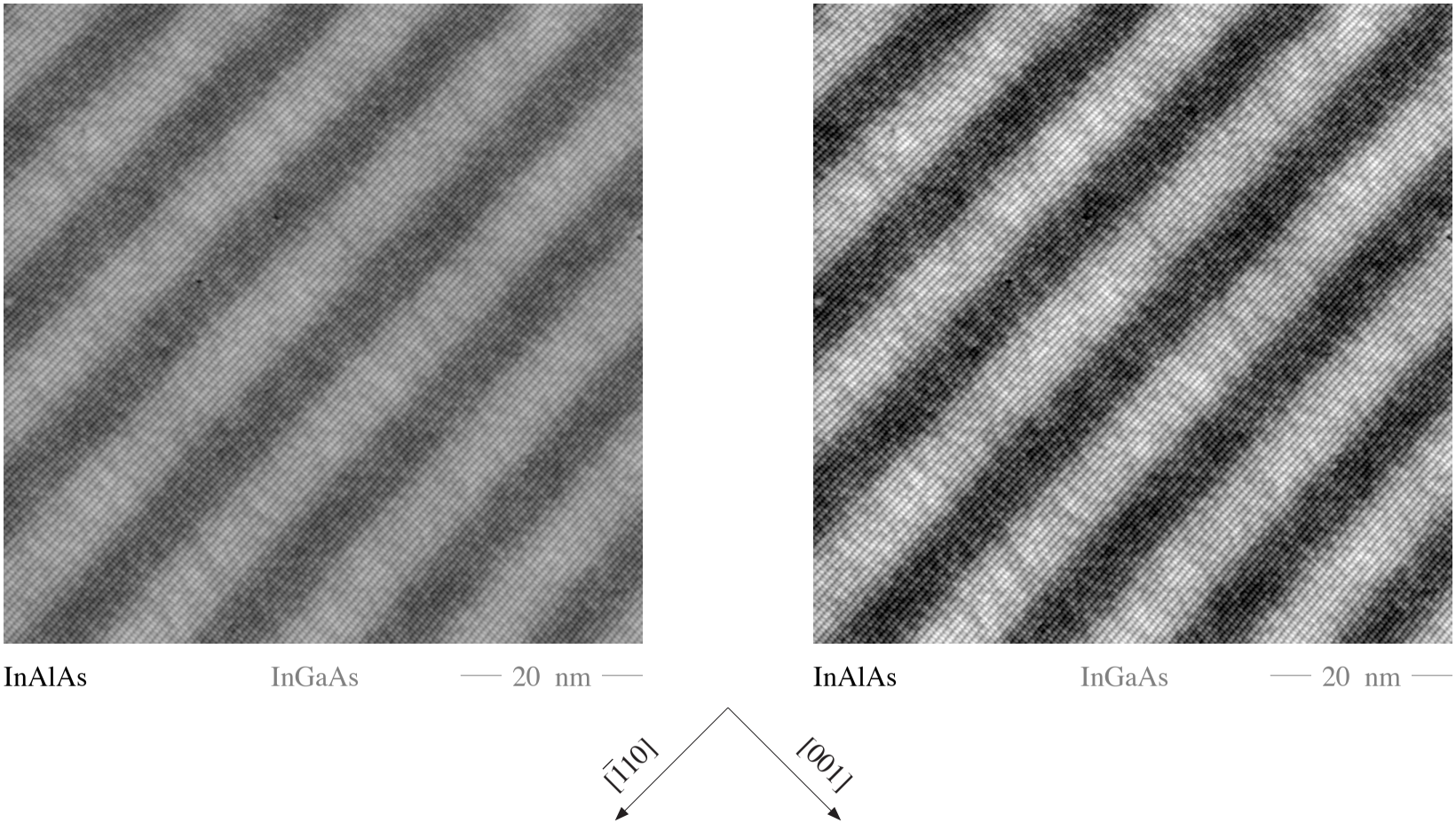


FIGURE 3.18. Native (left) versus contrast–enhanced (right) STM images from a strain–balanced InAlAs / InGaAs superlattice. Grey–level dispersion relative to the mean (height scale) is amplified by a factor of two (on the right) to increase layer contrast.

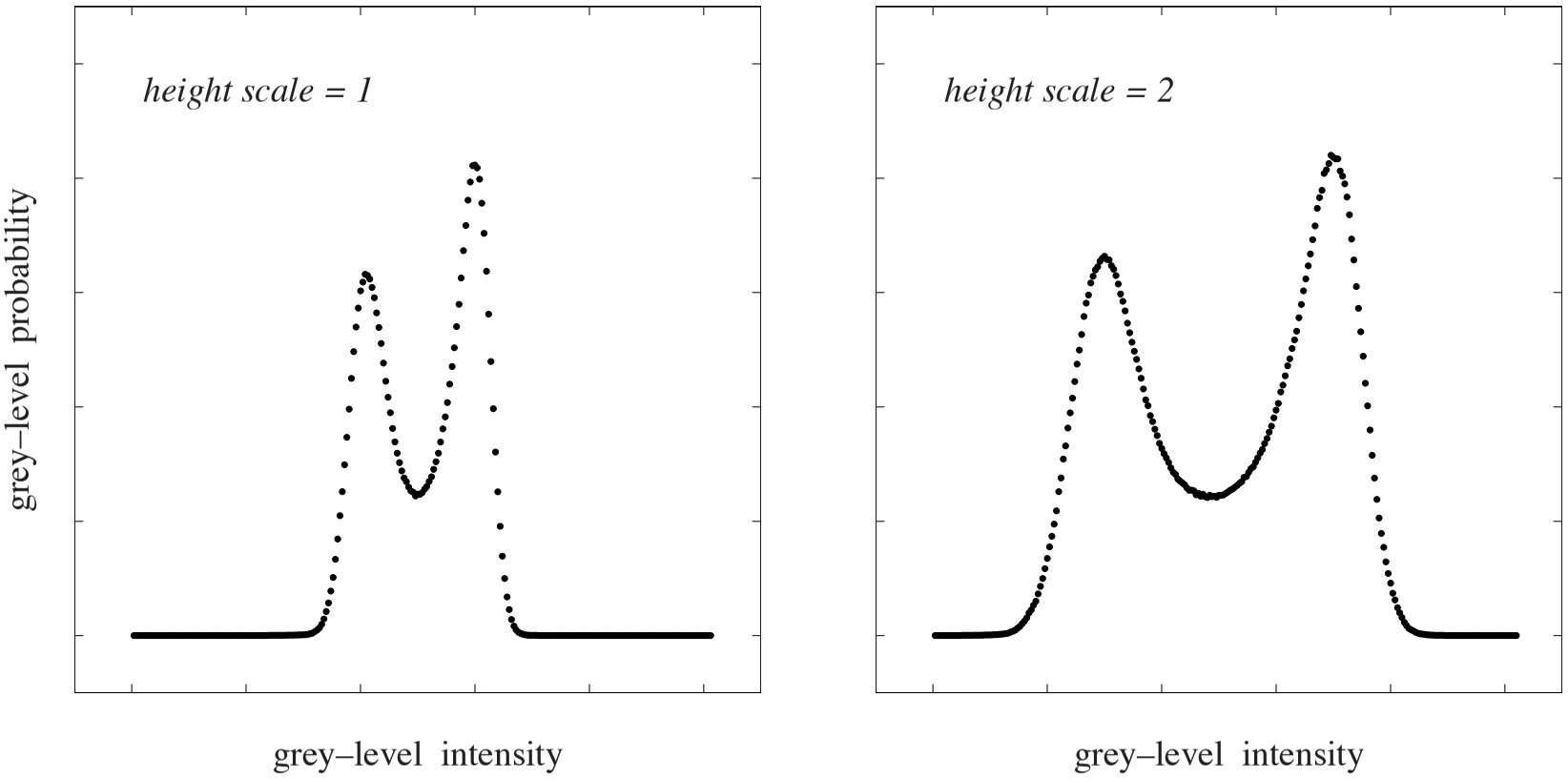


FIGURE 3.19. Bandwidth–limited, ensemble–average grey–level histograms based on native (left) versus contrast–enhanced height scale (right) with improved peak discrimination.

turned out much narrower than anticipated, our exploration of this parameter nevertheless provided several useful insights, described more fully below.

One, final, experimental parameter relates to an *a priori* conceivable dependence of the local state density imaged with STM on cleavage cross–section: anisotropic diffusion in the growth plane during epitaxial deposition might potentially influence the interface roughness amplitudes and / or correlation lengths observed in (110) versus (1–10) cross sections. The effect of cleavage cross–section on the contrast histograms subsequently used to generate heterojunction–delineating, binary masks is illustrated in Fig. 3.20. The data, here, are from the previously described surveys in Fig. 3.1, which were acquired from orthogonal cleaves over the same subset of superlattice periods and imaged with identical tunnel current and tunnel voltage. The distribution of grey levels in these complementary surveys is very similar, but they are also not exact copies of one other; in other words, even for surveys acquired with the same STM imaging conditions – albeit on different crystal facets – there are small, but quantifiable, differences in the detailed shape of the contrast histograms[4]. Indeed, the very fact these histograms do *not* share a universal shape was what fueled our curiosity concerning the role contrast threshold plays in any subsequent heterojunction delineation from binary masks (Figs. 3.14 and 3.15).

The effect of tunnel voltage on heterojunction delineation can be assessed – to the extent permitted by stable tunneling conditions – with the pair of surveys shown in Fig. 3.21. These surveys, previously introduced in Fig. 2.32 and acquired over the same

[4] Whether these (small) differences are crystallographic, or STM–tip related, is, unfortunately, something we are unable to convincingly address with such limited data.

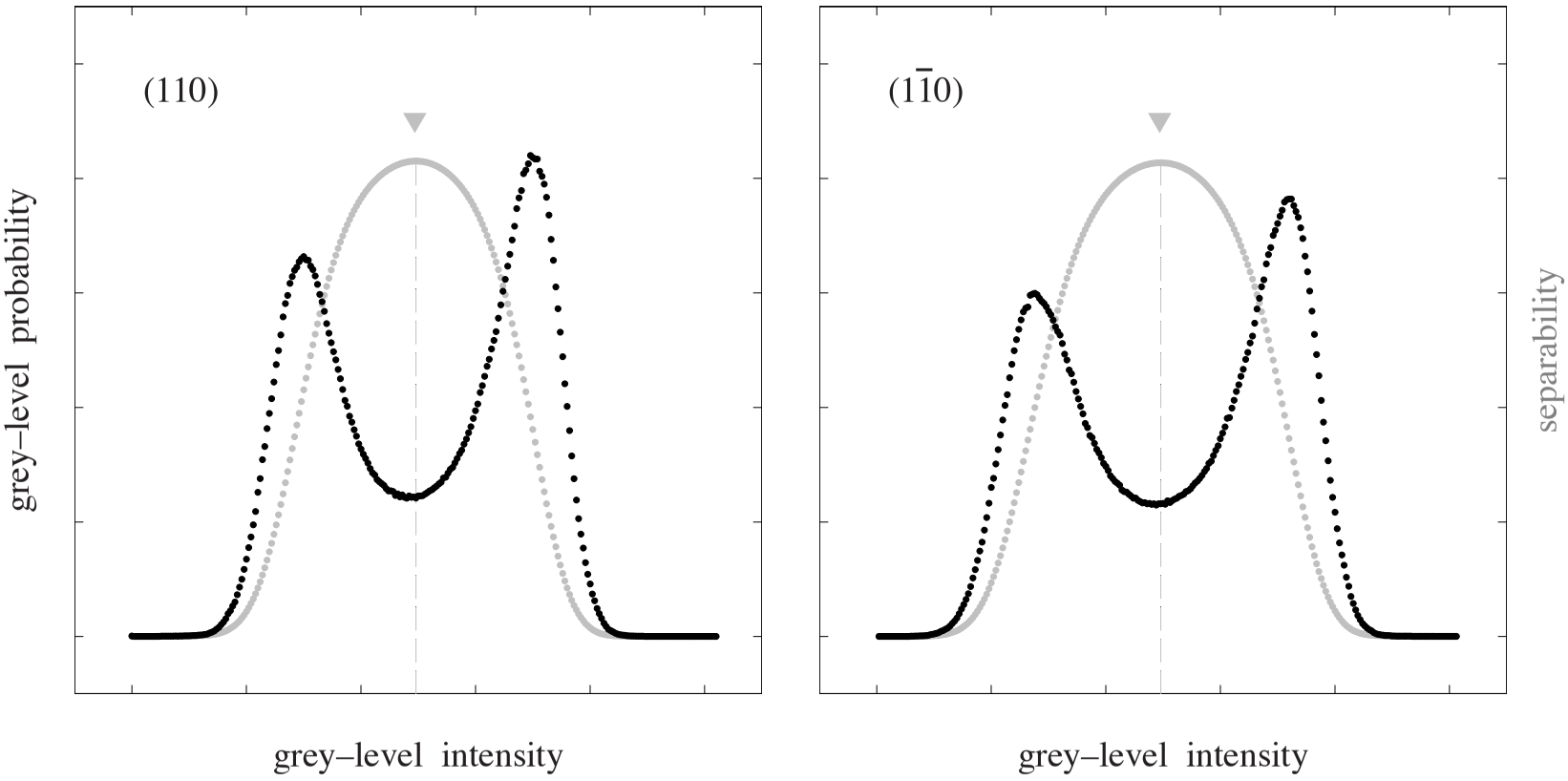


FIGURE 3.20. Bandwidth–limited, grey–level histograms from STM surveys over identical periods on neighboring dies (Fig. 2.3) exposed in (110) and (1–10) cross–sections, respectively. These histograms reflect qualitatively similar, but nevertheless distinguishable, grey–level distributions. Carat denotes Otsu threshold.

area on consecutive days, exhibit only minor differences in their respective grey–level histograms (Fig. 3.22), but binary masks constructed from the subset of survey images with near–perfect spatial overlap offer a more critical test of the effect of this (small) change in tunnel voltage on our heterojunction profiles. Two such sets of masks (one for each voltage) are overlaid in the respective panels of Fig. 3.23, and regions where the masks differ from one another are highlighted in red. These discrepancies are infrequent[5], and located primarily along the InAlAs–on–InGaAs interface. Although the one–quarter–volt sample bias difference examined here may appear comparatively small, in absolute terms, it nevertheless represents a significant fraction of the laboriously–determined window of tip stability, estimated to be, at best, only three times as large.

Heterojunction Extraction – Basics

Our paradigm for heterojunction discrimination has up till now focused on an essentially geometry–independent object – the Fermi–filtered contrast histogram – to define an equi–contrast contour that most effectively separates our periodic quantum wells from their confining barriers and vice versa. To subsequently extract roughness amplitudes and correlation lengths from the binary masks so obtained, however, we must briefly reconsider the several geometric non–idealities associated with STM data acquisition that occupied much of our attention in Chapter II.

[5] On the order of five–percent of the surveyed interfaces.

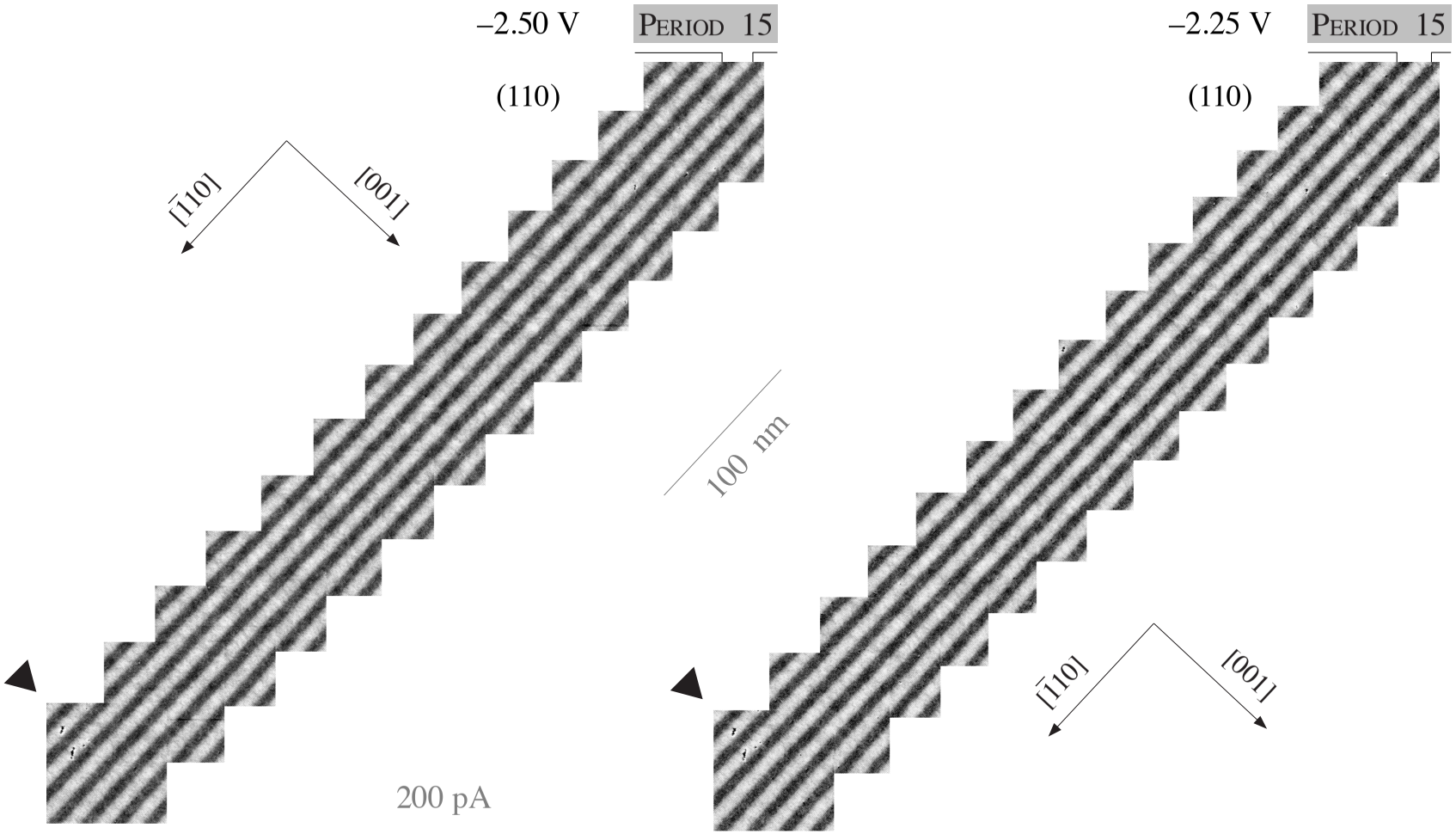


FIGURE 3.21. Device–scale, atomic–resolution surveys over nominally "identical" areas of a strain–balanced InAlAs / InGaAs superlattice acquired with two distinct tunnel voltages. Carat pinpoints a cleavage–related imperfection used to establish spatial registry of the two surveys. Growth direction is from top left to bottom right in each case.

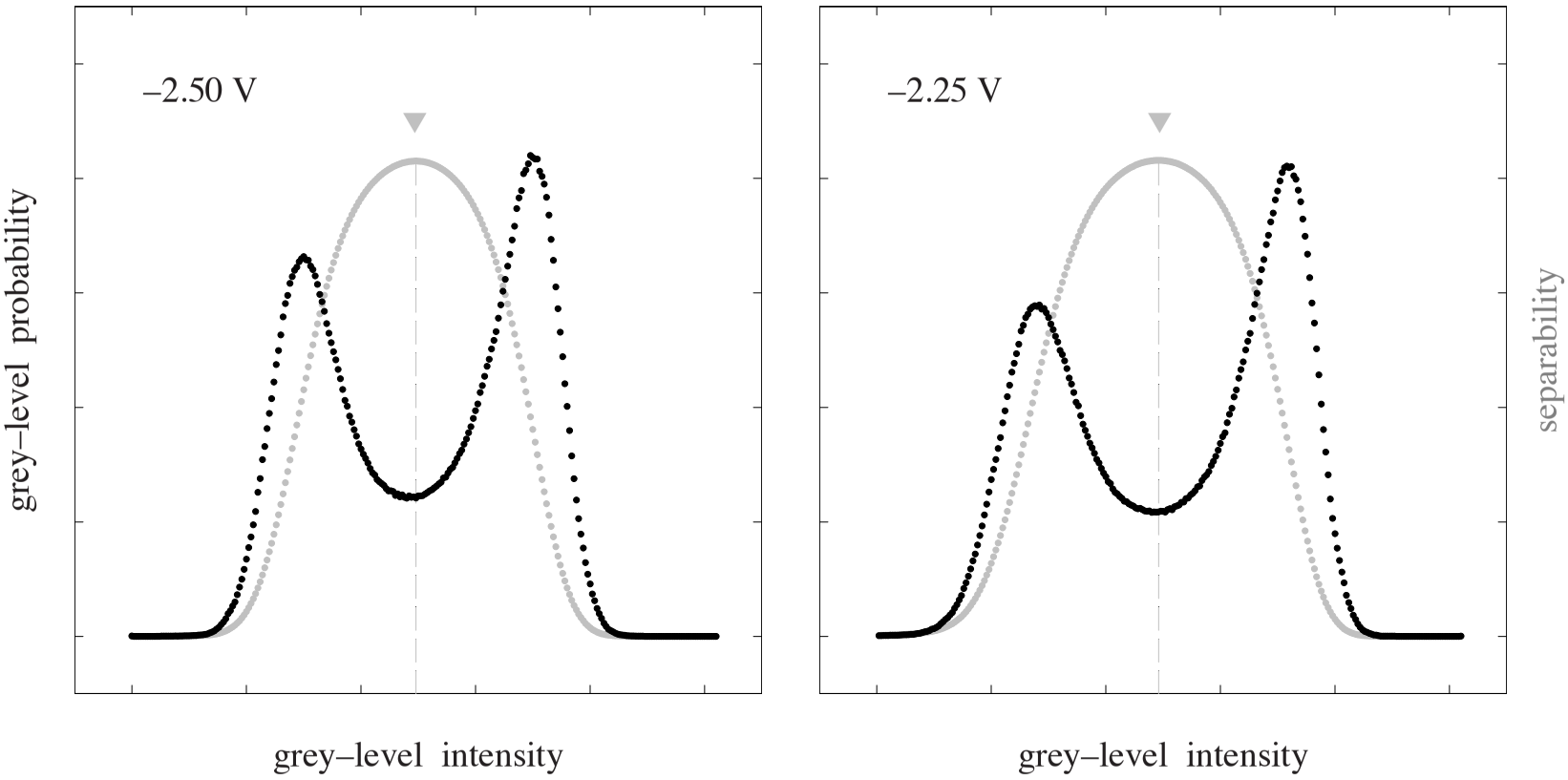


FIGURE 3.22. Bandwidth–limited grey–level histograms from the respective STM surveys in Fig. 3.21. These histograms, again, reflect very similar but distinguishable grey–level distributions. Carat denotes Otsu threshold.

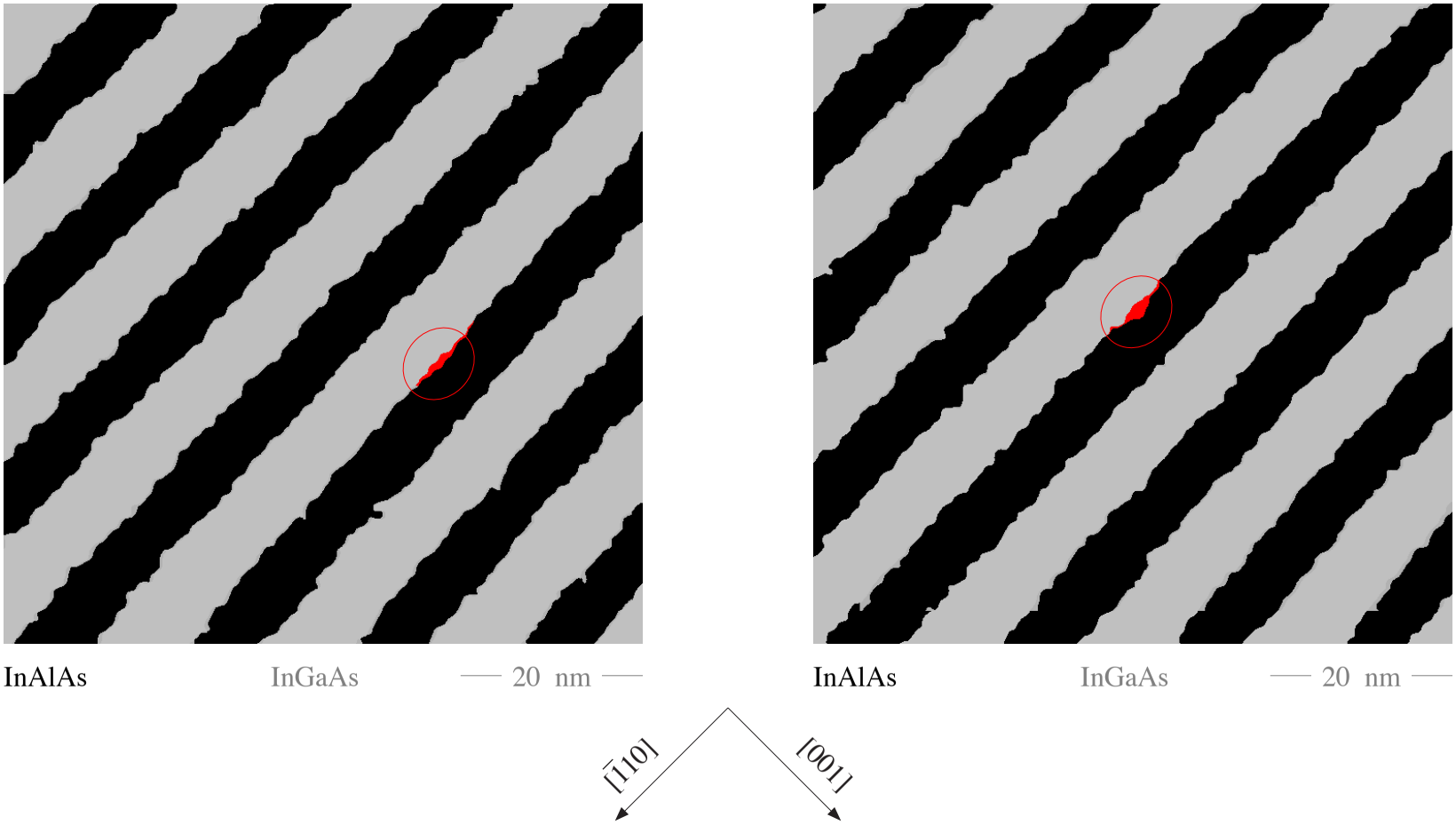


FIGURE 3.23. Influence of tunnel voltage on the binary masks from STM images with near–perfect spatial overlap. Each panel (left or right) compares a –2.50 V mask overlaid on its respective –2.25 V counterpart. Discrepancies between the two (highlighted in red) are infrequent ($\leq 5\%$), but, where they occur, are associated almost exclusively with the InAlAs–on–InGaAs interface. Growth direction is from top left to bottom right in each case.

The computation of reciprocal–space power spectra, from which correlation lengths will be derived in Chapter V, requires our spatial profiles represent a single–valued [001] function of a ⟨110⟩ growth–plane coordinate. As described above, this requirement is met by appropriate filtering that eliminates retrograde points. The needed pooling of individual, physically–equivalent power spectra into statistically–significant ensembles furthermore necessitates each profile comprise an equi–sampled sequence in this coordinate. This second condition can only be satisfied by rotating interfaces of nominally–identical length so that the ⟨110⟩ in–plane coordinate is oriented along the horizontal axis and subsequent digitization (of both this in–plane coordinate as well as fluctuations in the vertical growth direction) is performed in units of pixels, with normalization to a survey–average, in–plane reciprocal–lattice vector following (but not an image–by–image reciprocal–lattice vector prior to) pooling.

Pixel–based rotations, of course, assume the validity of a Euclidean geometry, as do all trigonometric calculations designed to determine an applicable rotation angle. This assumption may, at first, seem unwarranted in view of the piezo–scanner–related distortions described at length in Chapter II, but turns out to be entirely satisfactory, at the few percent level, for reasons we next explain. Within a suitably–conducted survey we have previously seen the *x*– (fast scan) component of b_{001} varies by no more than 0.5 %, whereas the *y*– (slow scan) component varies by as much as 3 – 4 %. The corresponding variation in $|b_{001}|$ is smaller than the latter by approximately a factor of two, however, on account of components being added in quadrature, and the same holds true for $|b_{110}|$; both magnitudes are invariant under pixel–based (orthogonal)

transformations.

A somewhat surprising, but equally important fact is the following: calculations of ensemble rms roughness conducted in either of two ways, first employing image–by–image normalization to $|b_{001}|$ and subsequently averaging over all images, or, alternatively, first ensemble averaging and subsequently normalizing to the survey–average $|b_{001}|$, differ from each other by at most a few parts in ten thousand.

We now describe the sequence of digital–processing steps necessary to isolate, extract, rotate and interpolate interface profiles from the binary masks obtained by way of histogram thresholding. The first step is to visually enhance layer discrimination – while retaining atomic resolution – by digitally overlaying a transparent binary mask atop the corresponding STM image as shown in Figs. 3.24 and 3.25 for (110) and (1–10) cross sections, respectively. This overlay is essential in order to pair each mask–delineated interface of interest with a nearby atomic trough that serves as a crystallographic reference for determining the rotation angle needed to align the respective heterojunction with an equi–sampled, horizontal axis. A representative example, highlighting two consecutive sets of InAlAs–on–InGaAs (blue) and InGaAs–on–InAlAs (plum) interfaces from which specific angles (with respect to the *x*–axis) can be found, is shown in Fig. 3.26, left. The corresponding (unrotated) heterojunctions shown in Fig. 3.26, right, were extracted directly from the overlaid binary mask via a traditional Laplacian edge–detection algorithm [52].

Due to the piezo non–idealities considered at length in Chapter II, the (001) planes in our (forward–scan) STM images are not parallel to one another. Consecutive

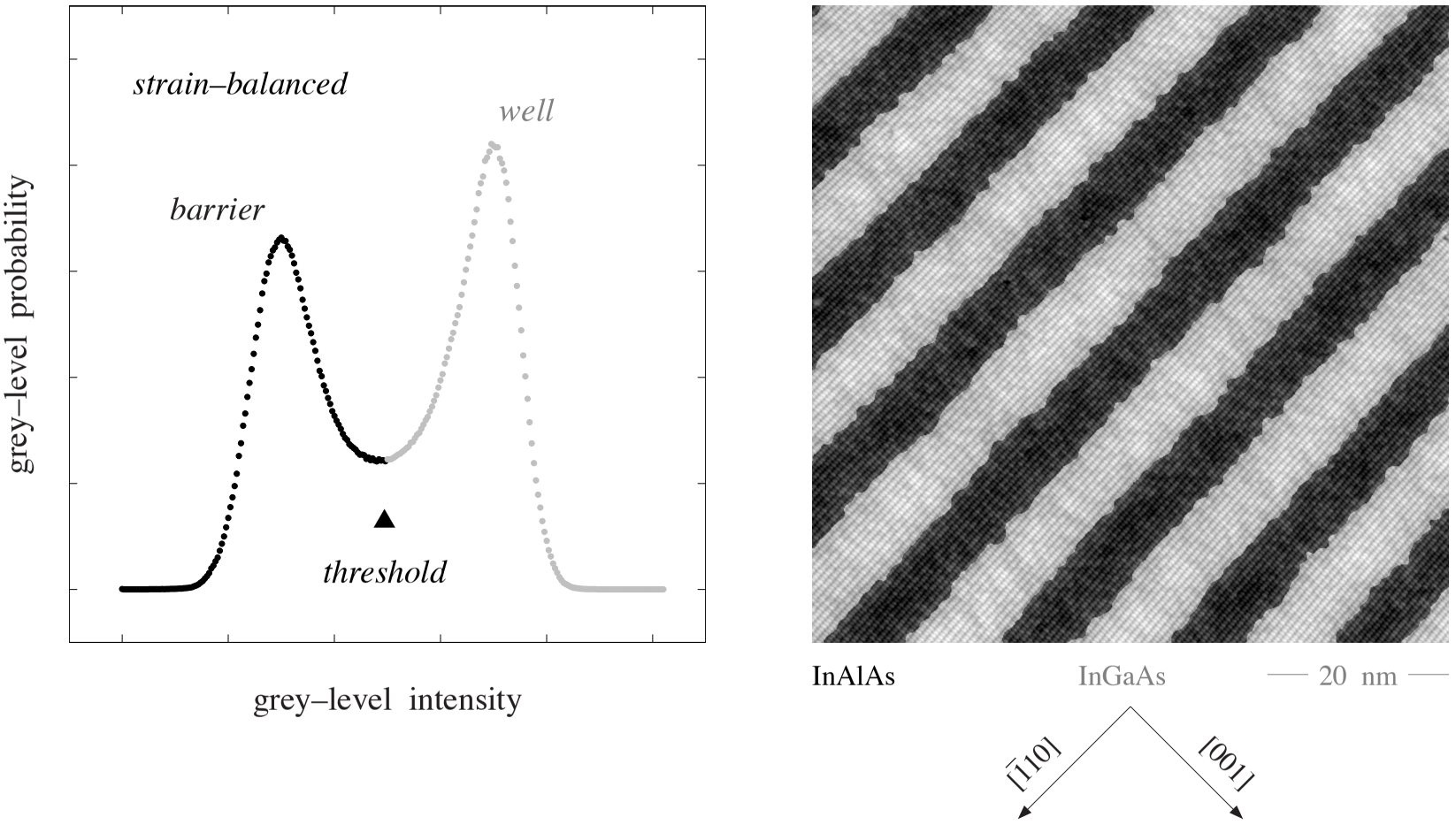
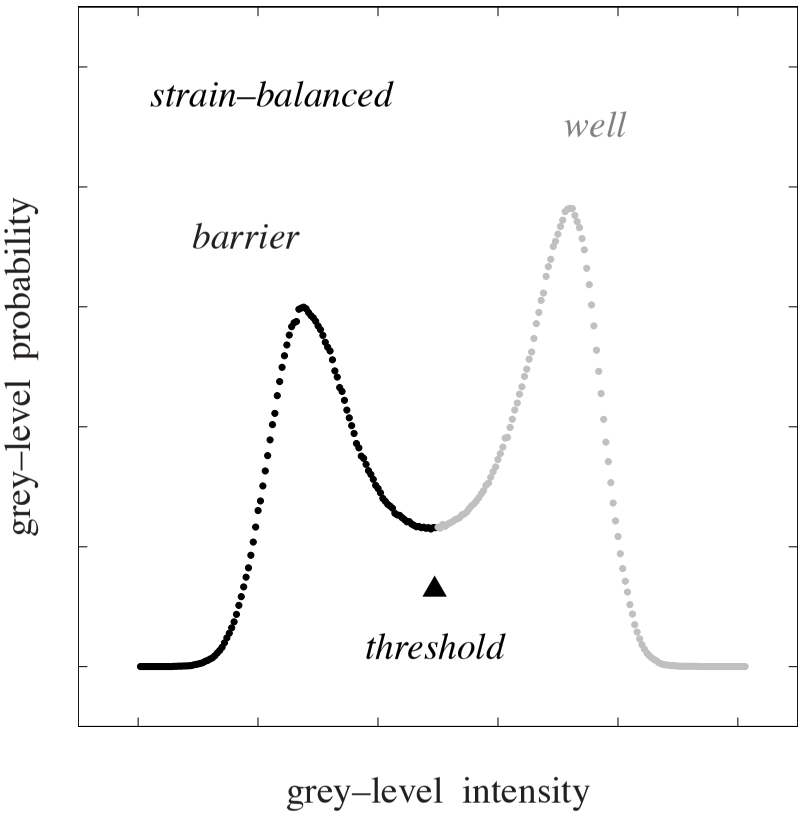


FIGURE 3.24. Bandwidth–limited contrast histogram (left) and atomic–resolution image (right) with superimposed binary mask to accentuate heterojunction delineation; STM data is from the (110) survey in Fig. 3.1 (left). Growth direction is from top left to bottom right.



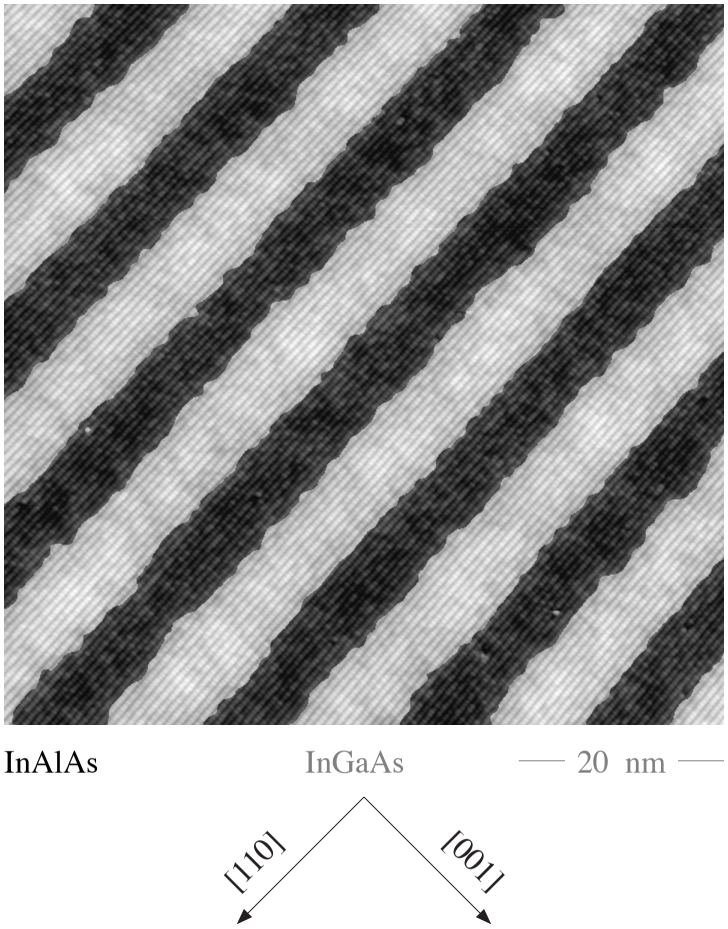


FIGURE 3.25. Same contrast histogram (left) and mask–enhanced image (right) as previously described in Fig. 3.24, but now in (1–10) cross section (Fig. 3.1, right).

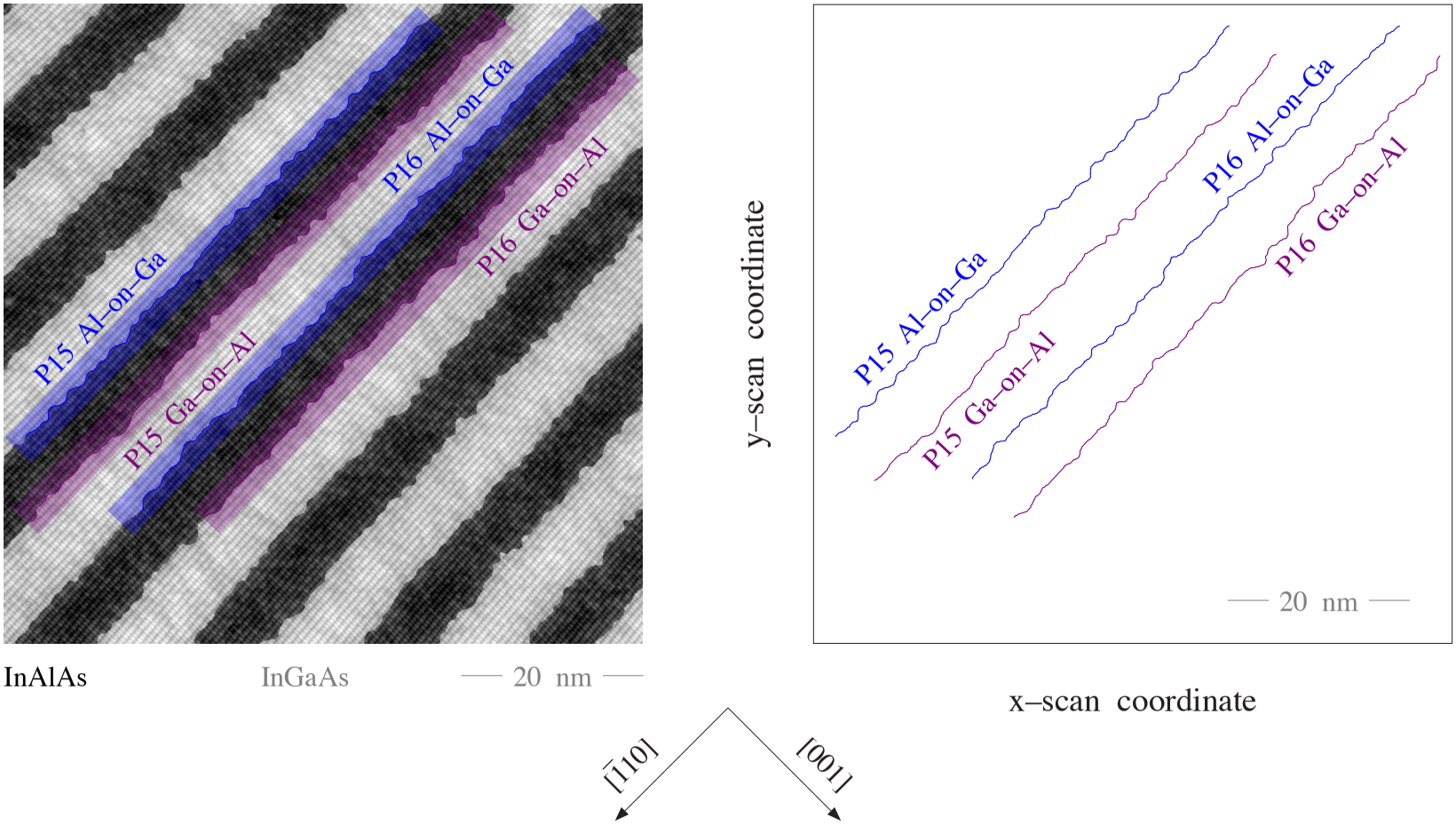


FIGURE 3.26. Contrast–enhanced, atomic–resolution STM image (left) highlighting four consecutive interface segments. Heterojunction profiles (right) are digitally extracted from the corresponding binary masks via a Laplacian edge–detection algorithm. Growth direction is top left to bottom right.

(001) planes threading the cleavage surface are, in fact, mis–oriented by about three–hundredths of a degree[6], leading to a net, accumulated mis–orientation of (roughly) one–third degree between consecutive interfaces. Since the inclination of each heterojunction with respect to the horizontal varies across the image, this angle must be individually and laboriously determined for each interface of interest. To further complicate matters, the angle of any given interface varies significantly over a survey as indicated in Fig. 3.27, right, where a typical image–by–image drift is illustrated with four consecutive interfaces.

The angles summarized in Fig. 3.27 (right) are obtained from the slope of a straight line traced along the closest atomic trough paralleling each heterojunction [54]. This labor–intensive step would not be necessary if the piezoelectric tripod was perfectly orthogonal and linear, since an "average" interface angle can efficiently be determined from the image power spectrum. The 2–D DFT of each STM image is computed from the shaded area indicated in Fig. 3.28 (left) whose diagonal lies very close to P15 Ga–on–Al interface. As seen in Fig. 3.28 (right), the "average" angle obtained this way from the corresponding b_{001} reciprocal–lattice vector agrees remarkably well with the real–space value independently assigned to that interface. Fig. 3.29 (right) illustrates the effect of subsequent rotation, through appropriate angles determined from Fig. 3.27, on the four, digitized interfaces considered in Fig. 3.26 (right).

The reliance on atomically–resolved crystallographic troughs to determine interface angles illustrates one further complication generally anticipated with epitaxial

[6] a complication further exacerbated in reverse–scan images

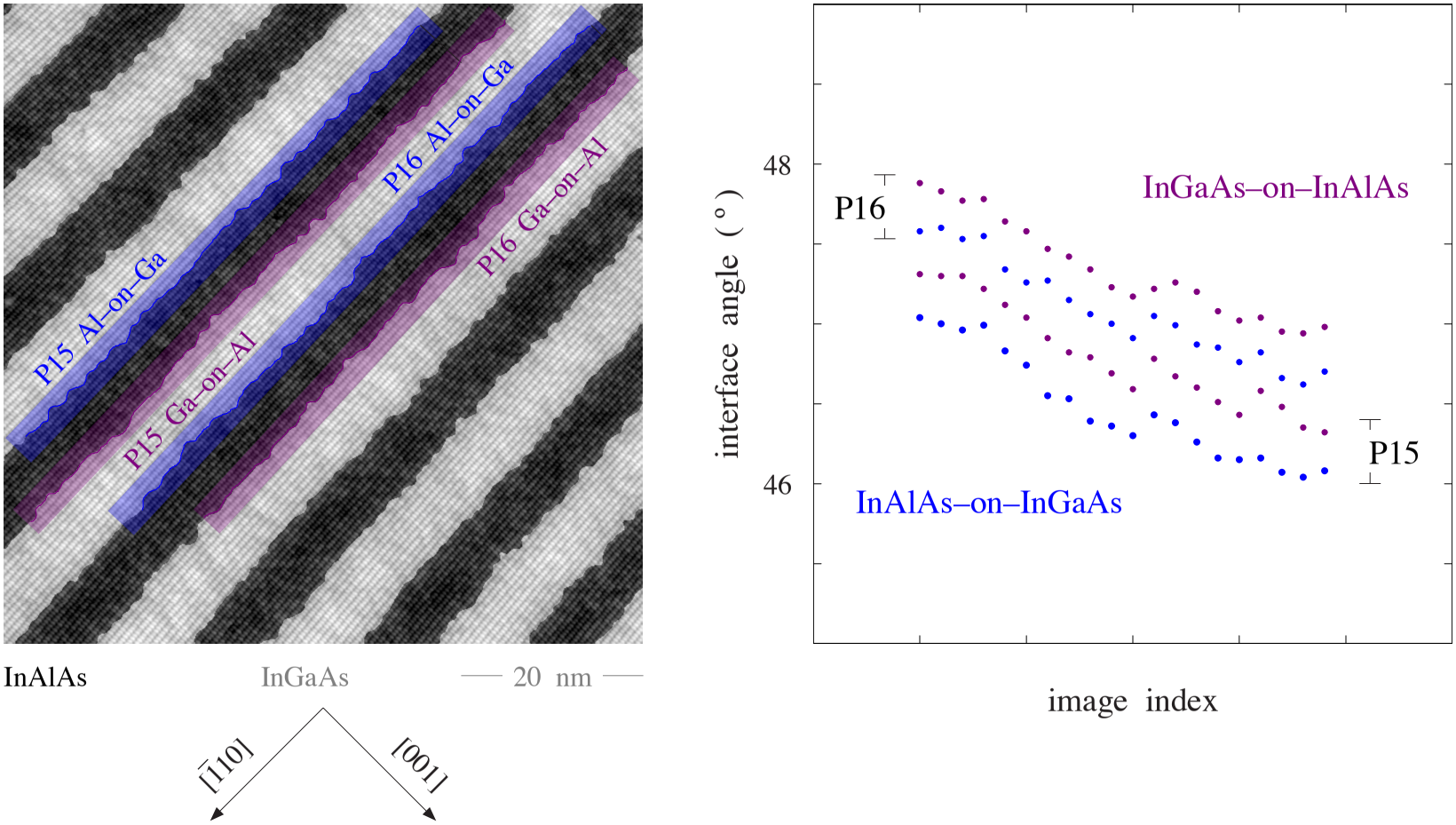


FIGURE 3.27. Contrast–enhanced, atomic–resolution STM image (left) and highlighted interface angles (right) needed for subsequent heterojunction profile rotation. Due to remnant distortion, consecutive interfaces within a given image are misaligned by ~ 0.3°, and the apparent inclination of one interface varies by as much as one degree over a complete survey.

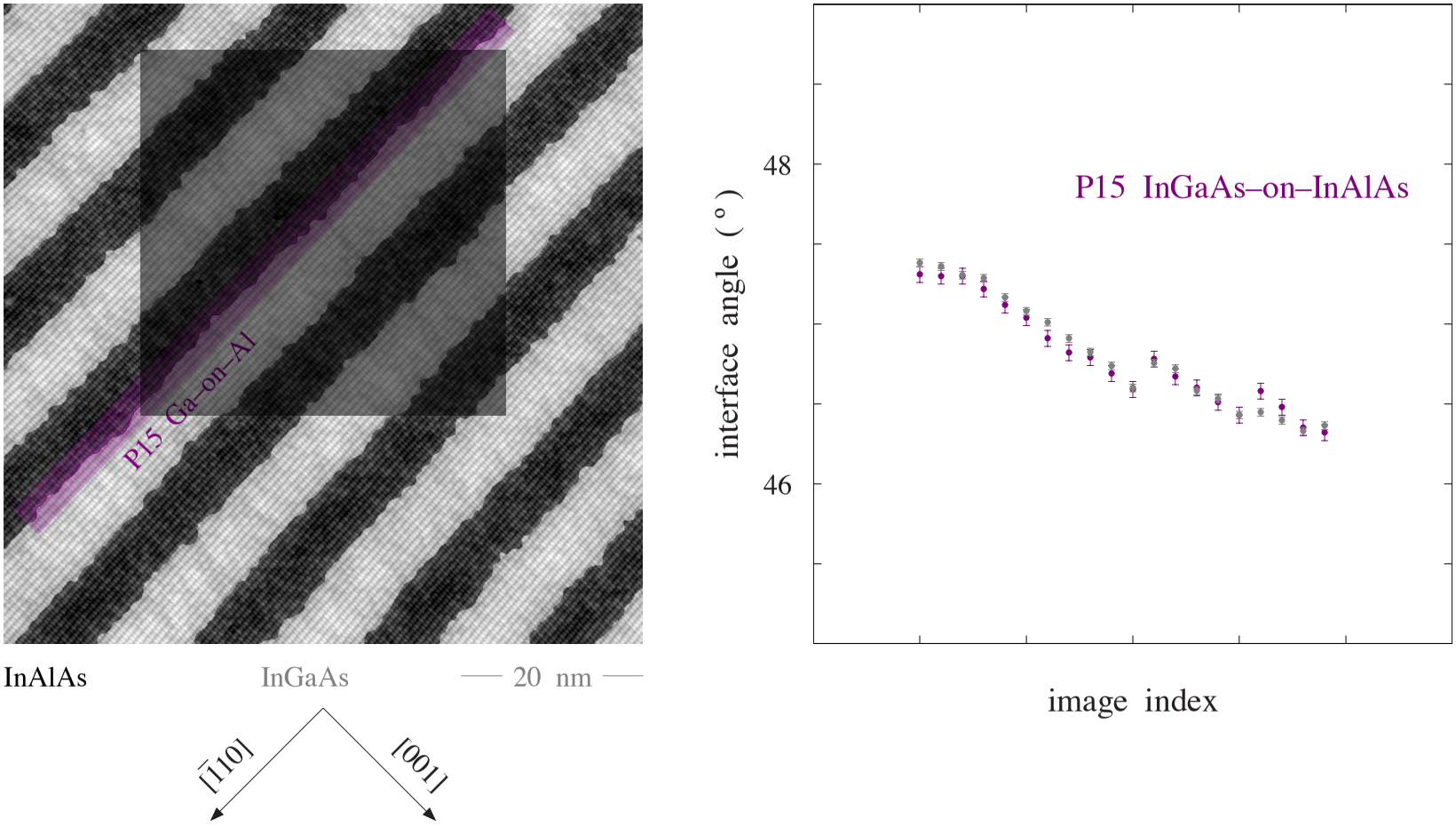


FIGURE 3.28. Contrast–enhanced, atomic–resolution STM image (left) with shaded overlay indicating the square region used for DFT computation. The "average" interface angle deduced from a DFT–generated reciprocal–space map (right, grey) agrees with the corresponding real–space angle for the main diagonal (right, plum) from Fig. 3.26.

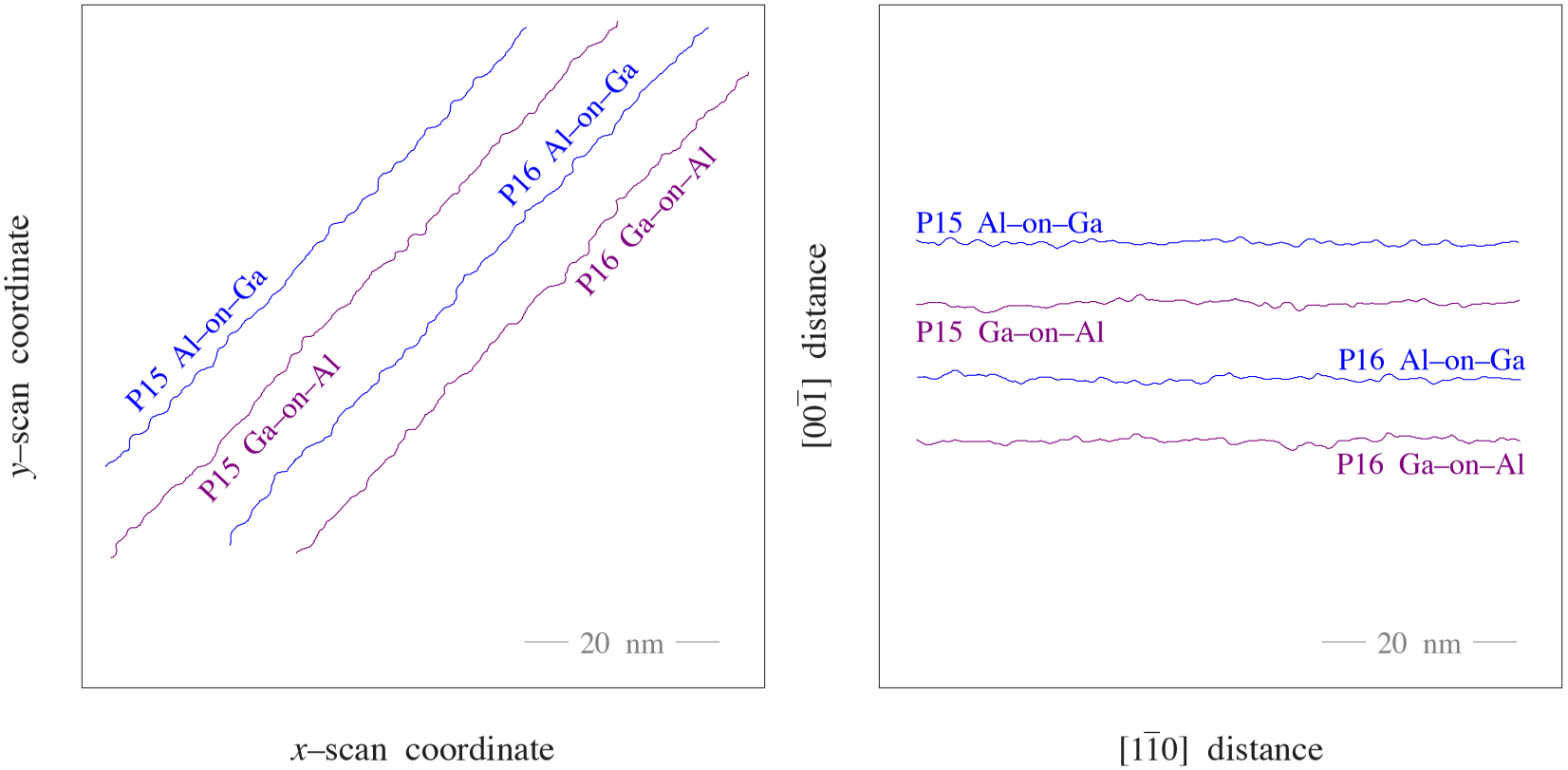


FIGURE 3.29. Effect of rotation on digitally–extracted interface profiles. Original orientation (left) does not guarantee the equidistant sampling along ⟨110⟩ needed for power spectrum computation because of the discrete character of STM raster scans; rotated profiles (right) do, since the relevant crystallographic axes are now aligned with *x*– and *y*–scan increments, respectively.

growth. Any substrate wafer that serves as a growth template will exhibit some degree of geometrical miscut with respect with to an ideal (001) plane; this miscut angle (≤ 0.1° for the substrate wafers here) is specified as an isotropic tolerance, but its actual value can be different in the two, linearly–independent ⟨110⟩ growth–plane directions. Construction of (110) and (1–10) survey mosaics from carefully–aligned STM images allows direct measurement of this inclination. As illustrated in (110) cross section with Fig. 3.30, initially–nearby troughs and interfaces gradually loose their registry so that, towards the conclusion of a micron–long survey, they diverge from one another by about two surface monolayers. This divergence reflects an approximate 0.08º misalignment with respect to the [1–10] primitive translation vector; a slightly smaller misalignment (0.06° with respect to the [110] direction) is similarly found in (1–10) cross section. Since our tabulated interface angles are tied to crystallographic atomic troughs, this (admittedly small) residual angular bias (or "vicinality") should be removed from all profiles prior to analysis.

Heterojunction Extraction – Revisited

There is one final complication stemming from piezo–raster non–idealities that must now be addressed: the angular inclination of any given ⟨110⟩ atomic trough with respect to the horizontal – what, in effect, will become our power spectral density abscissa following rotation – not only evolves in the [001] direction, as already described above, but changes, ever so slightly, along the [110] (or [1–10]) as well. This "bowing" of (001) planes with respect to the straight line best "globally" aligned with a

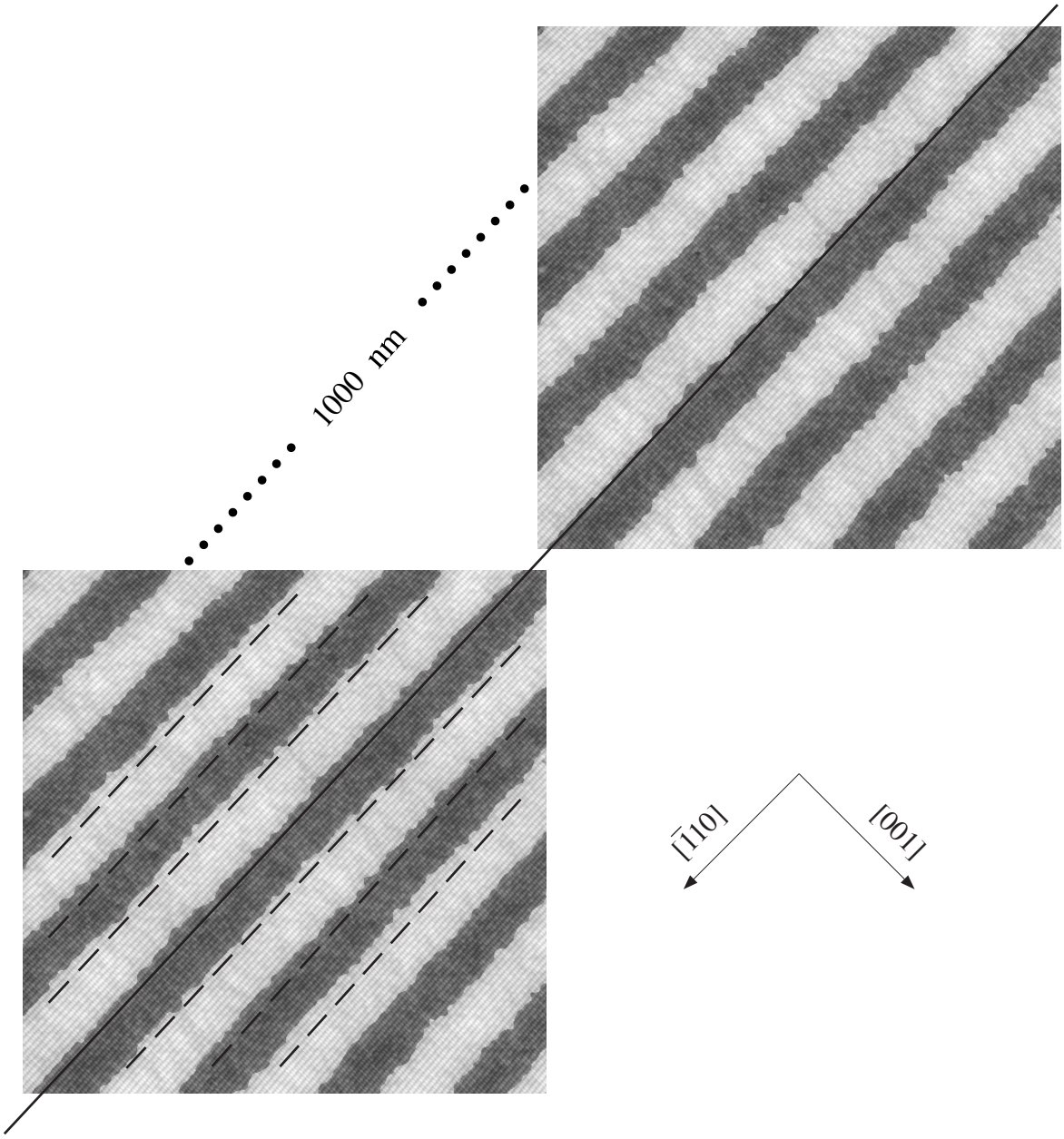


FIGURE 3.30. Determination of substrate vicinality from micron–long lateral surveys. Careful alignment of consecutive images throughout a survey mosaic facilitates accurate measurement of any accumulated deviation between superlattice interfaces (embedded within a tilted growth plane) and nearby atomic troughs (solid line) that serve as fixed, crystallographic references parallel to (001) lattice planes (dashed lines).

nearby atomic trough is illustrated in Fig. 3.31.

Fig. 3.31 presumes, as is in fact the case with the angular measurements detailed in the preceding section, that this best "global" alignment is achieved with a line anchored at the end points of the particular interface segment in question. To lowest order then, [001] "bowing", if uniform in nature, will be described by a parabola symmetrically positioned about the midpoint of a global ⟨110⟩ chord as illustrated in Fig. 3.31 (left). Fig. 3.31 (right) illustrates the ensuing consequence of this assumption for our data; if the chord in Fig. 3.31 (left) is bisected, and the original interface segment thereby cut into two, equal–length halves, there will be an angular "parallax" (or divergence) equal to the difference in half–segment inclinations and proportional to the original chord length. The appropriateness of such a model for our STM images is motivated with an explicit example in Fig. 3.32.

Figs. 3.33 and 3.34 present two alternative strategies for carving up an STM image into identifiable interface segments suitable for further analysis. Each interface is labeled by the corresponding superlattice period and heterojunction growth order. Despite every conceivable refinement of experimental technique (including carefully executed re–scans of the affected area) severely non–uniform bowing irreparably distorts the geometry of the lower eighth of any STM image, leaving the data in this region unusable for present purposes. The partitioning suggested in Fig. 3.33 is based on a premise of maximum interface length – consistent with the rectangular boundaries of our STM images – within the minimally–distorted, upper seven–eighths region of the scan; the heterojunction segments illustrated here range between 350 and 775 Å in length, in

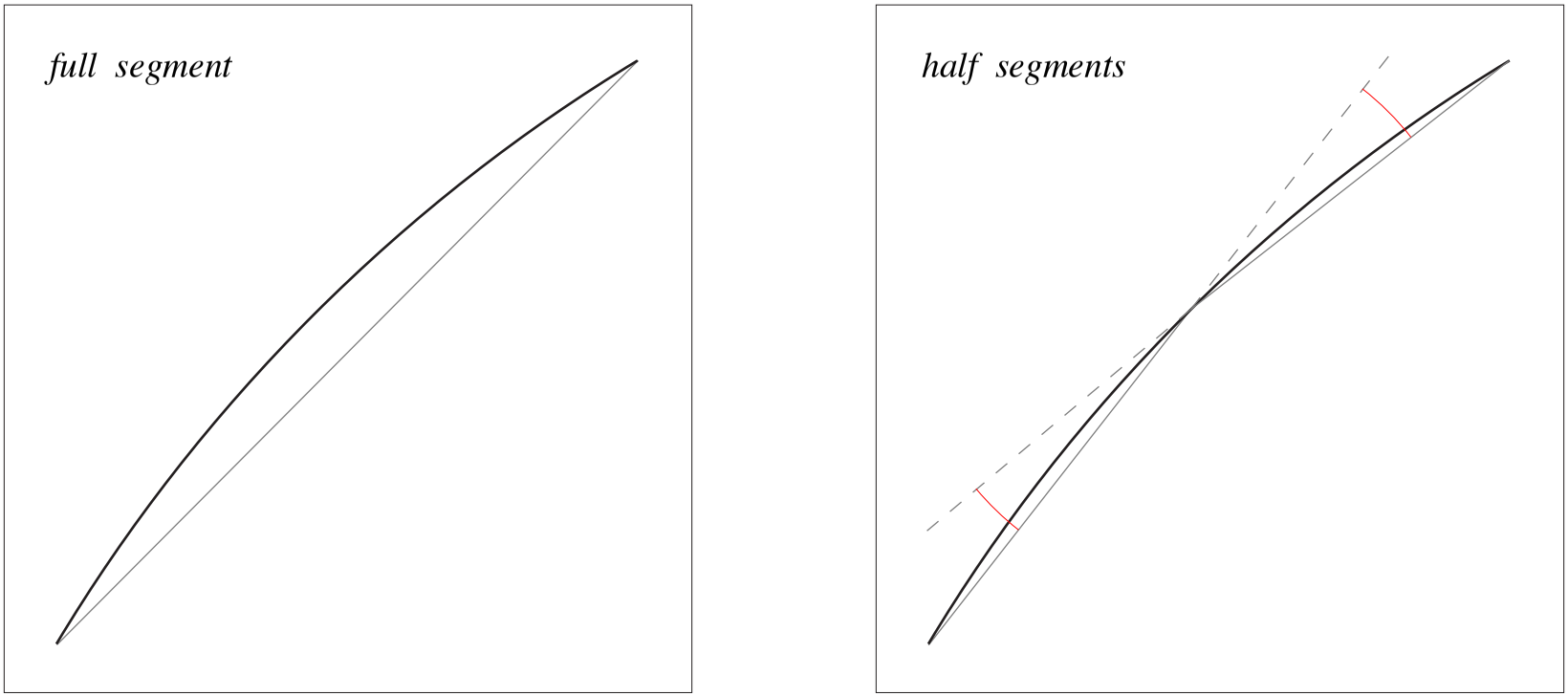


FIGURE 3.31. Schematic illustration of STM–raster–induced distortion of (001) lattice planes. Curved, black line represents a parabolically–bowed (001) plane and the grey chord its best "global" linear approximation. This growth–plane bowing is quantified via the angular divergence (red), or parallax, between half segments whose projections (dashed lines) accentuate the misalignment between chord endpoints.

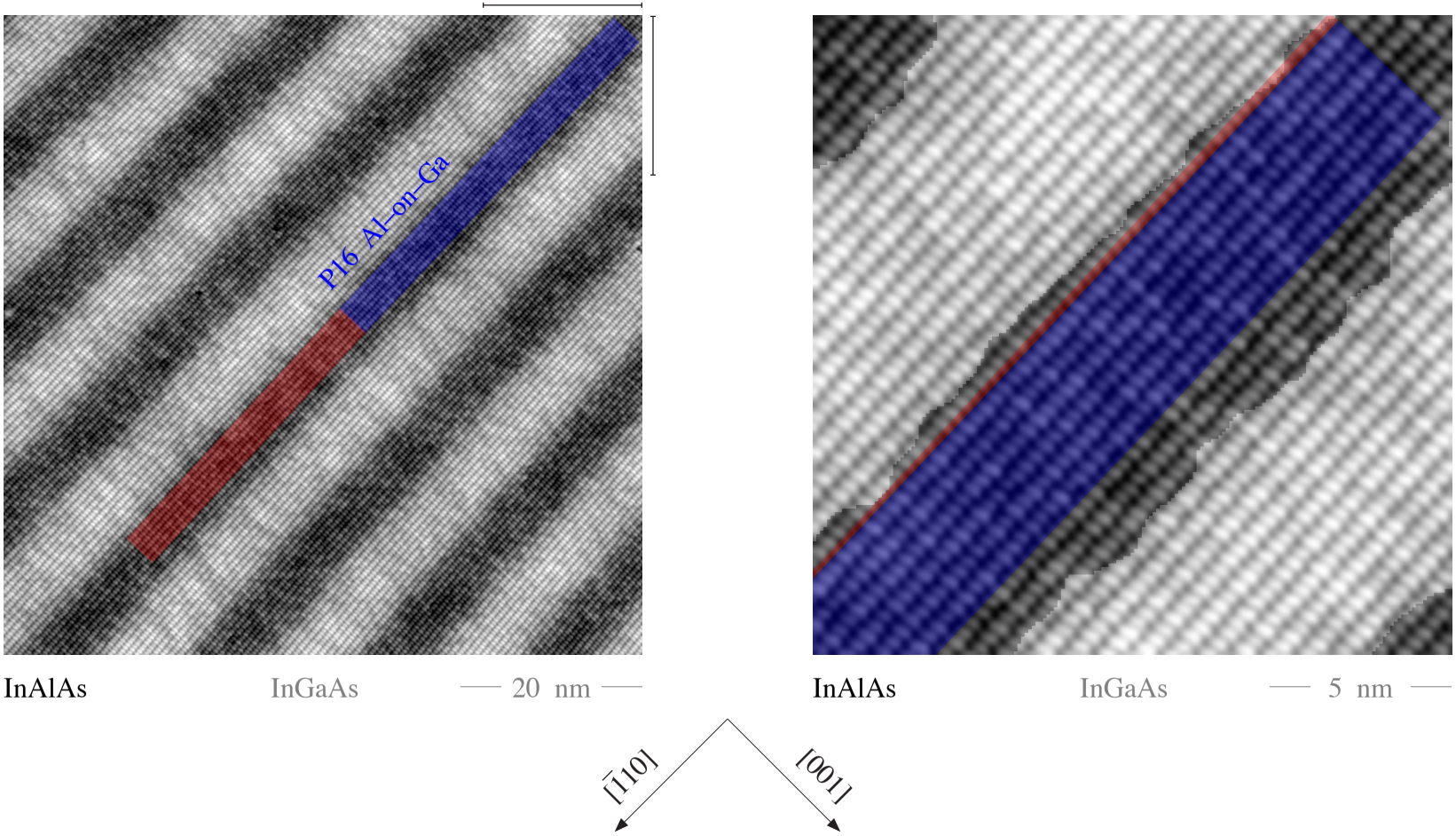


FIGURE 3.32. Schematic construction of Fig. 3.31 applied to actual STM data; red and blue rectangles (left) are locally aligned with the their respective half–segment interface chords. Projection of the red rectangular chord into the upper right region of the STM image illustrates the monolayer–scale angular parallax purposely exaggerated in Fig. 3.31 (right).

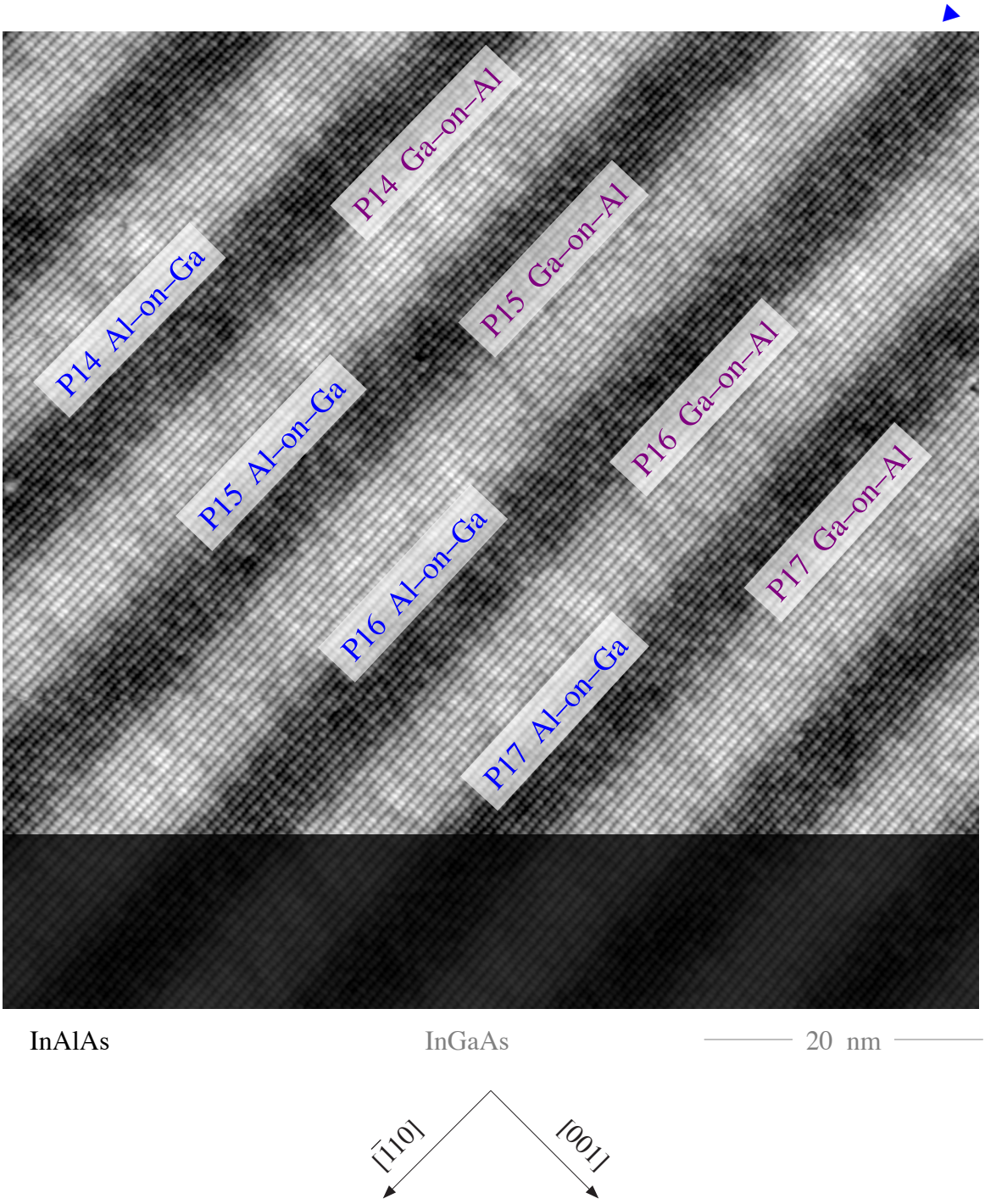


FIGURE 3.33. STM image partitioning that maximizes individual interface–profile lengths. Shaded rectangular is omitted due to non–uniform curvature accompanying the initiation of a raster scan. Caret points to a reference interface that may be followed in subsequent figures. Growth direction is top left to bottom right.

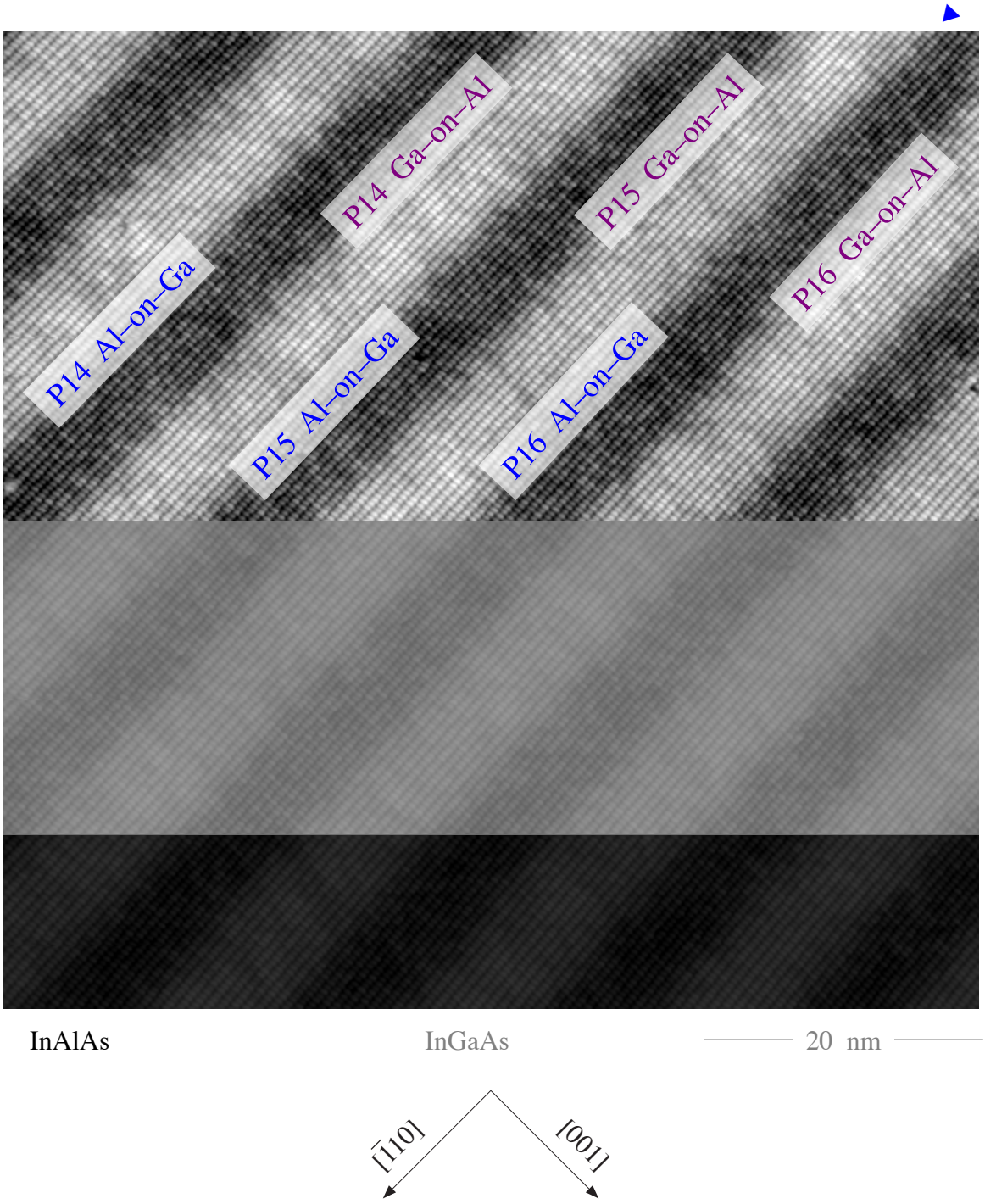


FIGURE 3.34. Refined partitioning of Fig. 3.33 designed to enforce a common interface–segment length to preserve the option of pooling power spectra from individual interface profiles into statistical ensembles.

principle facilitating the detection of correspondingly–long correlation lengths. Fig. 3.34 begins from an altogether different perspective: insist on uniform segment lengths to maximize the number of individual power spectra available for pooling into physically–equivalent, statistical ensembles. The interface segments here are all 350 Å in length.

Fig. 3.35 (left) contrasts the typical, mid–survey, angular parallax attending the variable–length–segment strategy indicated in Fig. 3.33 with the fixed–length strategy of Fig. 3.34 now applied to a featureless, homogeneous substrate. We immediately learn two things from this comparison: first, the bowing of fixed–length substrate segments varies in relation to their location within the STM image; second, the bowing of similarly–placed superlattice segments additionally varies in direct relation to heterojunction length. Fig. 3.35 (right) demonstrates that when the parallax in Fig. 3.32 is normalized to a common 350Å segment length, superlattice and substrate bowing are indistinguishable. This linear scaling with segment length is a clear indication the observed bowing is, in fact, uniform in the (restricted) sense described above.

We consider, next, the influence of this scan–related curvature of (001) lattice planes on the experimental quantities of central interest to us here: mean roughness variance and the roughness power spectrum. The respective biases introduced by bowing of fixed– and variable–length segments into the standard computation of mean–square segment deviations are contrasted in Fig. 3.36; the predicted offset is, in all cases, comparatively small (as results in Chapter IV indicate[7]), but shorter, uniform–length segments are undeniably preferable. Corresponding biases in the (roughness normalized)

[7] c.f. table in page 178 for comparison.

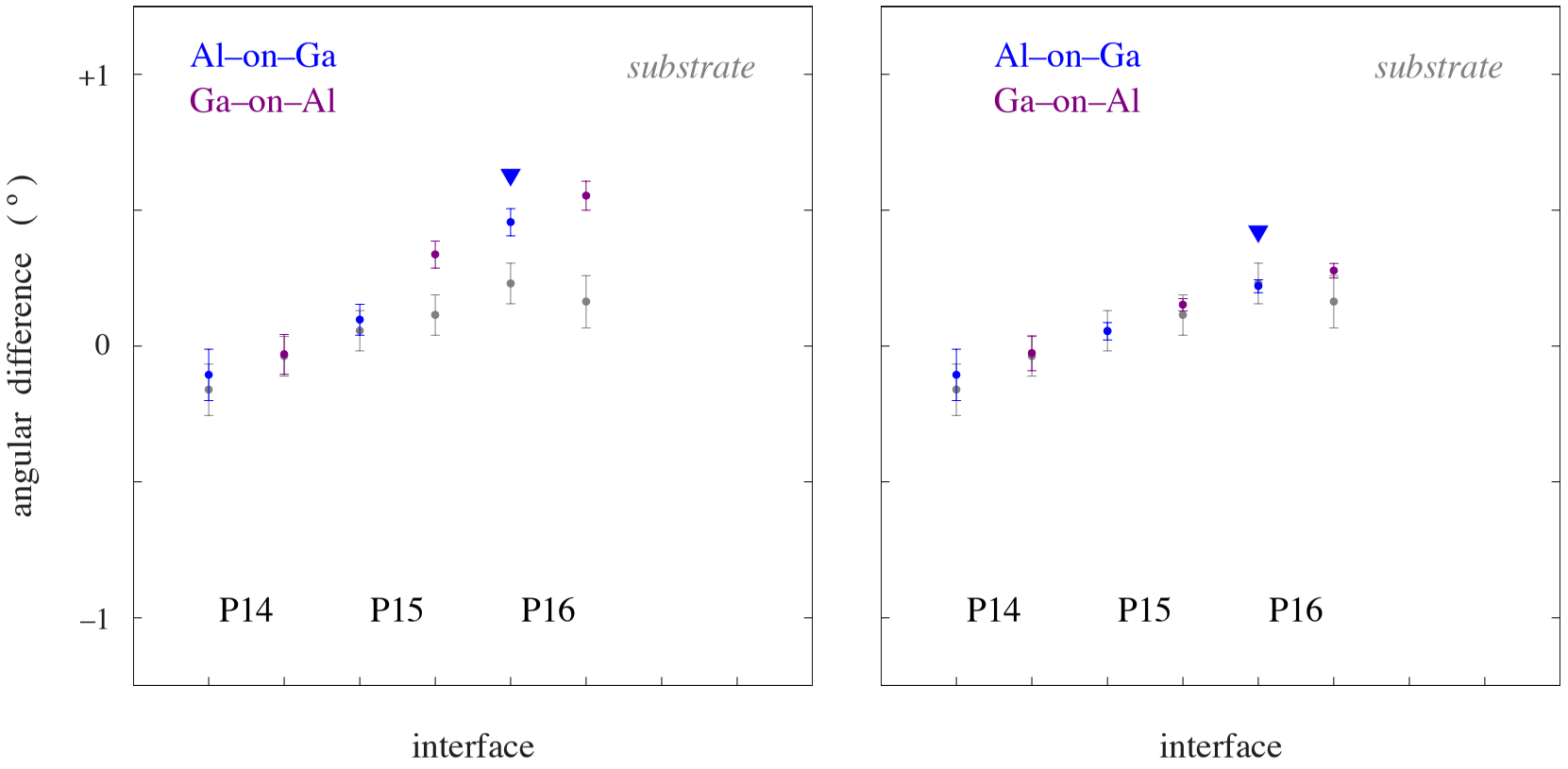


FIGURE 3.35. Angular divergence between variable–length half–segments (per Fig. 3.33) contrasted with similar measurements from shorter, uniform–length segments (per Fig. 3.34) on a featureless substrate, before (left) and after (right) normalization to a common 350 Å length. An evident change in sign with period (left) underscores the scan–frame position dependence of the noted curvature; the evident linear scaling with segment length (right), on the other hand, emphasizes its uniformity with respect to arc length along a given interface.

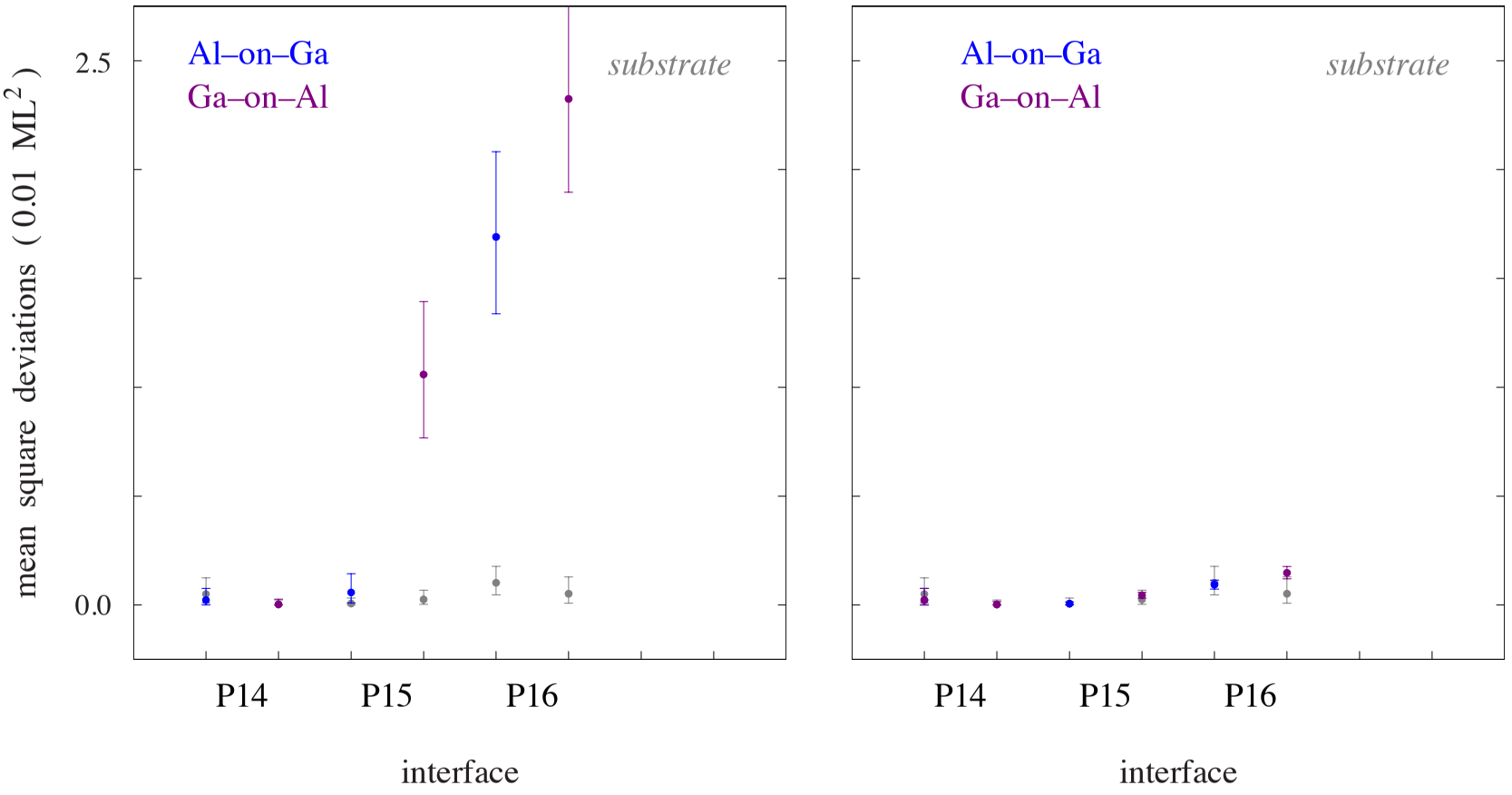


FIGURE 3.36. Effect of parabolic distortion of experimental heterojunction profiles on the roughness variance by using angular differences typical from large (colored-left) and small (grey) crop areas as those presented in Fig. 3.37, left. The effect of parabolic distortion from the large crop area (colored–left) normalized to a common 350 Å length (colored–right) agrees with the parabolic distortion influence for the featureless substrate measured over the shortened 350 Å.

power spectrum are indicated in Fig. 3.37, where 350 Å segments – with bowing parallax of 0.25° – and 750 Å segments – with bowing parallax 0.50° – are directly compared; the associated power spectra evidence a distinctive, inverse–fourth–power fall off whose influence on the experimental spectrum will be proportional to the ratio of the variance in Fig. 3.36 to the expected roughness variance (~ 1 ML), i.e., likewise small.

A cautious, but by no means compelling, conclusion to be drawn from Figs. 3.36 and 3.37 is that Fig. 3.34 is the better partitioning strategy for our STM images on geometrical grounds. The decisive argument in favor of Fig. 3.34 versus 3.33 is not geometry, however, but the flexibility only equal–length interface segments afford with respect to eventual grouping of their power spectra into statistical ensembles.

Preliminary Grouping of Heterojunction Profiles into Statistical Ensembles

The well–understood growth–plane anisotropy of semiconductor surface reconstructions [55] and monolayer islanding [56], together with direct evidence for growth–order–dependent asymmetries in the compositional intermixing at representative, epitaxially–formed III–V semiconductor interfaces [57] suggest from the very beginning that one may expect both cleavage orientation and interface type to be important physical variables for the development (and description) of interface roughness.

With this general backdrop in mind, it makes sense, absent compelling data to the contrary, to classify our experimental interface profiles into three, physically–distinct,

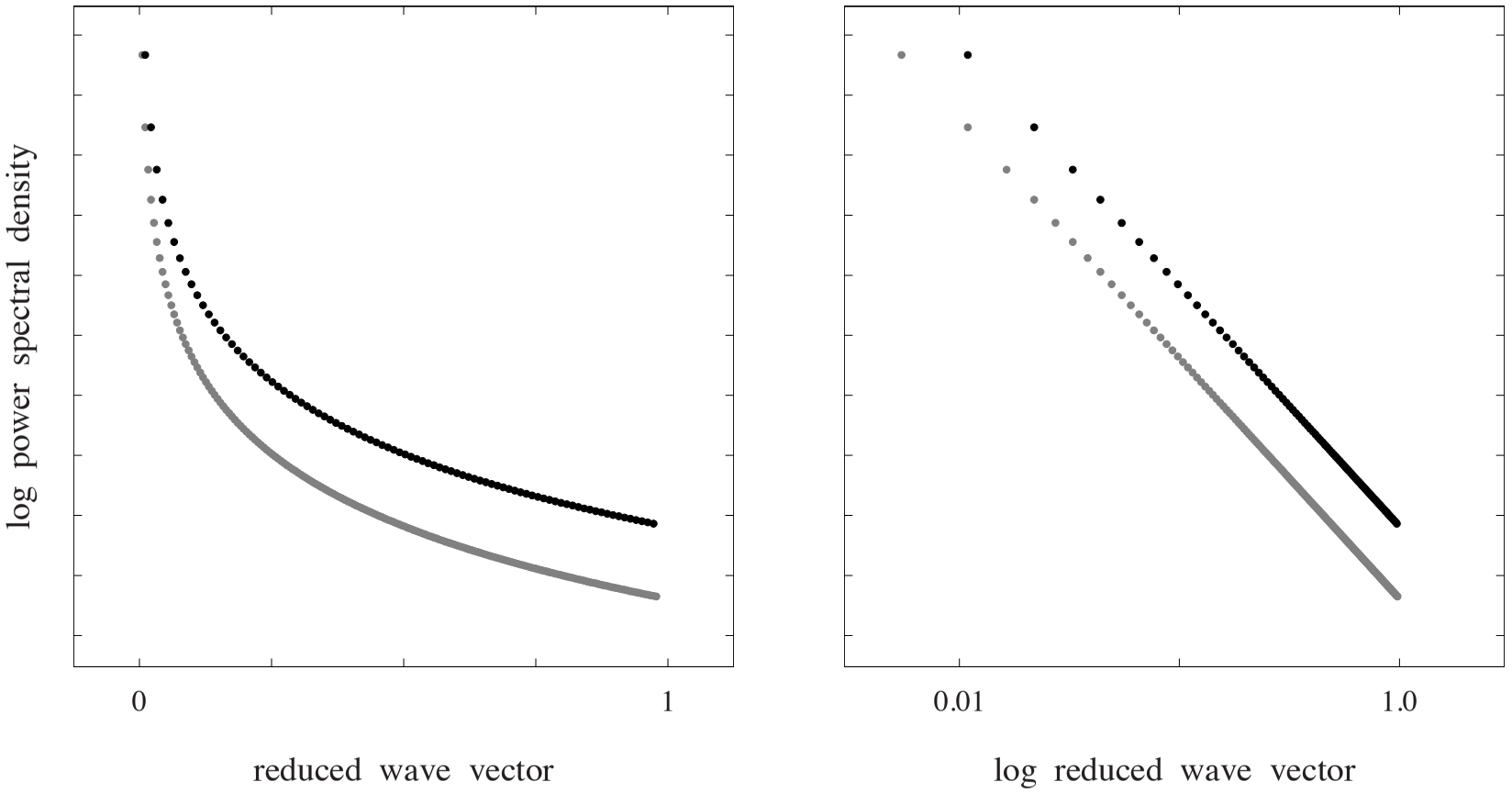


FIGURE 3.37. Normalized power spectral density (left) associated with uniform parabolic bowing of 350 Å– (black) and 700 Å–long (grey) interface segments and angular divergences (Fig. 3.31, right) of 0.25° and 0.50° respectively. Log–log plot (right) emphasizes the distinctive inverse–fourth–power dependence on spatial frequency.

statistical ensembles distinguished by heterojunction type – specifically, InAlAs–on–InGaAs (or Ga–template) vs InGaAs–on–InAlAs (or Al template) – cleavage–exposed cross section – (110) vs (1–10), and tunnel junction parameters (–2.50 V vs –2.25 V). Table 3.1 summarizes the number of digitized interface segments – period resolved as well as period pooled – retained in each ensemble following visual inspection of the original STM images[8].

We assume, from here on, that period–pooling is beyond question, but note the (period–pooled) physically–motivated ensembles in Chapter V[9] can also be subsequently pooled to provide better statistics should later analyses indicate they may, indeed, be legitimately combined.

[8] 102 – out of the total 336 – heterojunction profiles were rejected primarily due to imperfections near or along the interface such as adsorbates, tip–induced noise, and retrograde points.
[9] c.f. tables in pages 191 and 200.

Table 3.1. Heterojunction ensemble classification breakdown (grey) by survey, superlattice period, and interface template. Number of images taken as well as those either kept or rejected is indicated. Our period–pooled ensembles (black) used for subsequent analyses are likewise indicated.

		(110)			$(1\bar{1}0)$			(110)		
		–2.50 V, 20 images			–2.50 V, 16 images			–2.25 V, 20 images		
period	template	kept	rejected	% retained	kept	rejected	% retained	kept	rejected	% retained
14	Ga	11	9	55 %	12	4	75 %	9	11	45 %
	Al	12	8	60 %	12	4	75 %	15	5	75 %
15	Ga	14	6	70 %	15	1	94 %	14	6	70 %
	Al	14	6	70 %	14	2	88 %	14	6	70 %
16	Ga	15	5	75 %	9	7	56 %	12	8	60 %
	Al	17	3	85 %	8	8	50 %	17	3	85 %
14–16	Ga	40	20	67 %	36	12	75 %	35	25	58 %
	Al	43	17	72 %	34	14	71 %	46	14	77 %

CHAPTER IV

ROUGHNESS AMPLITUDE CHARACTERIZATION

Introduction

Optical transitions in QC lasers are engineered by means of quantum confinement in the conduction band; the heterojunctions adjoining the quantum wells and barriers are assumed to be planar, however the interface profiles presented in Chapter III call into question that assumption. Deviations from planarity can, in principle, adversely influence device properties, with energy broadening and carrier scattering being two common unintended consequences; to judge the strength of this influence we must quantitatively assess, both, the characteristic roughness amplitude and the lateral correlation lengths which together describe the stochastic processes governing the experimentally acquired interface profiles.

This chapter focuses on the analytical subtleties associated with the computation of the mean–square deviations, or squared roughness amplitude, of experimentally–acquired heterojunction profiles; complementary extraction of correlation lengths from the power spectral densities of these profiles is explored in the following chapter. Both analyses are needed to fully characterize the statistical properties of interface fluctuations, meaningfully reflect on their possible physical origins, and construct the scattering matrix elements [12] needed to correctly predict energy–level broadening and carrier lifetimes in QC active regions.

Experimental Profiles: Reconstruction of Fluctuation Amplitude Distributions

Computation of a mean roughness amplitude for any one of our experimentally–defined heterojunction ensembles is a seemingly straightforward task once these interfaces have been properly rotated and digitized. As we shall see in the immediately following section, where both the accuracy and the statistical certainty of such measurements is systematically analyzed with the help of numerical simulations, this question is considerably more subtle (and hence more complex) than generally acknowledged due to the very correlations we seek to understand.

Our study, here, begins with an analysis of the roughness amplitudes characterizing interface profiles extracted from the STM surveys in Fig. 3.1. The corresponding data ensembles distinguish two, potentially–different, interface types based on growth order – InAlAs–on–InGaAs versus InGaAs–on–InAlAs – as well as two, potentially–distinct subpopulations based on cross section – (110) versus (1–10) – whose physical properties might differ on account of growth–plane anisotropy, while averaging the data from periods 14 through 16. This organizational structure preserves four, physically–distinct sets of data that, if warranted, might then be further combined to improve statistics. An additional pair of interface–resolved ensembles emerges from separate consideration of the role played by imaging voltage versus cross section (as explored with the pair of STM surveys in Fig. 3.21), bringing our total number to six.

There are essentially two ways to gauge the roughness amplitudes associated with these heterojunction profiles. The first is straightforward tabulation of the variance per interface segment (relative to the corresponding segment mean) and calculation of

the corresponding ensemble average. This treatment, which focuses on a single number, discards potentially important information concerning the statistical distribution and functional dependence of the vertical deviations; a more insightful approach relies on the inferences drawn from the underlying probability distributions characterizing these observations for each experimental ensemble.

It is essential to point out, here, that tracking the absolute (i.e., consistently–referenced) coordinates for any one interface throughout an entire STM survey appears infeasible as a practical matter. For a host of reasons, many of which have already been described in Chapter III, the simplest, most reliable choice is to reference the fluctuations along any one interfacial segment to the corresponding segment mean. As we will see shortly, this seemingly innocuous – but unavoidable – decision carries profound consequences for the accuracy of naively–calculated roughness amplitudes, together with their statistical uncertainties, if these data are inherently correlated as expected.

Figs. 4.1 and 4.2 present the sampled distribution of vertical displacements from period–pooled, InAlAs–on–InGaAs (left) and InGaAs–on–InAlAs (right) interfaces on (110) and (1–10) cross sections, respectively, following heterojunction extraction, rotation, and digitization as previously detailed in Chapter III[1]. Fig. 4.3 presents the same data as Fig. 4.1, but obtained with –2.25 V versus –2.50 V sample bias.

These histograms are simply constructed by tabulating the number of occurrences of a given vertical excursion, $z_i - \langle z \rangle_j$, versus excursion amplitude – where i

[1] Including normalization to the magnitude of the appropriate, survey–averaged [001] reciprocal–lattice vector.

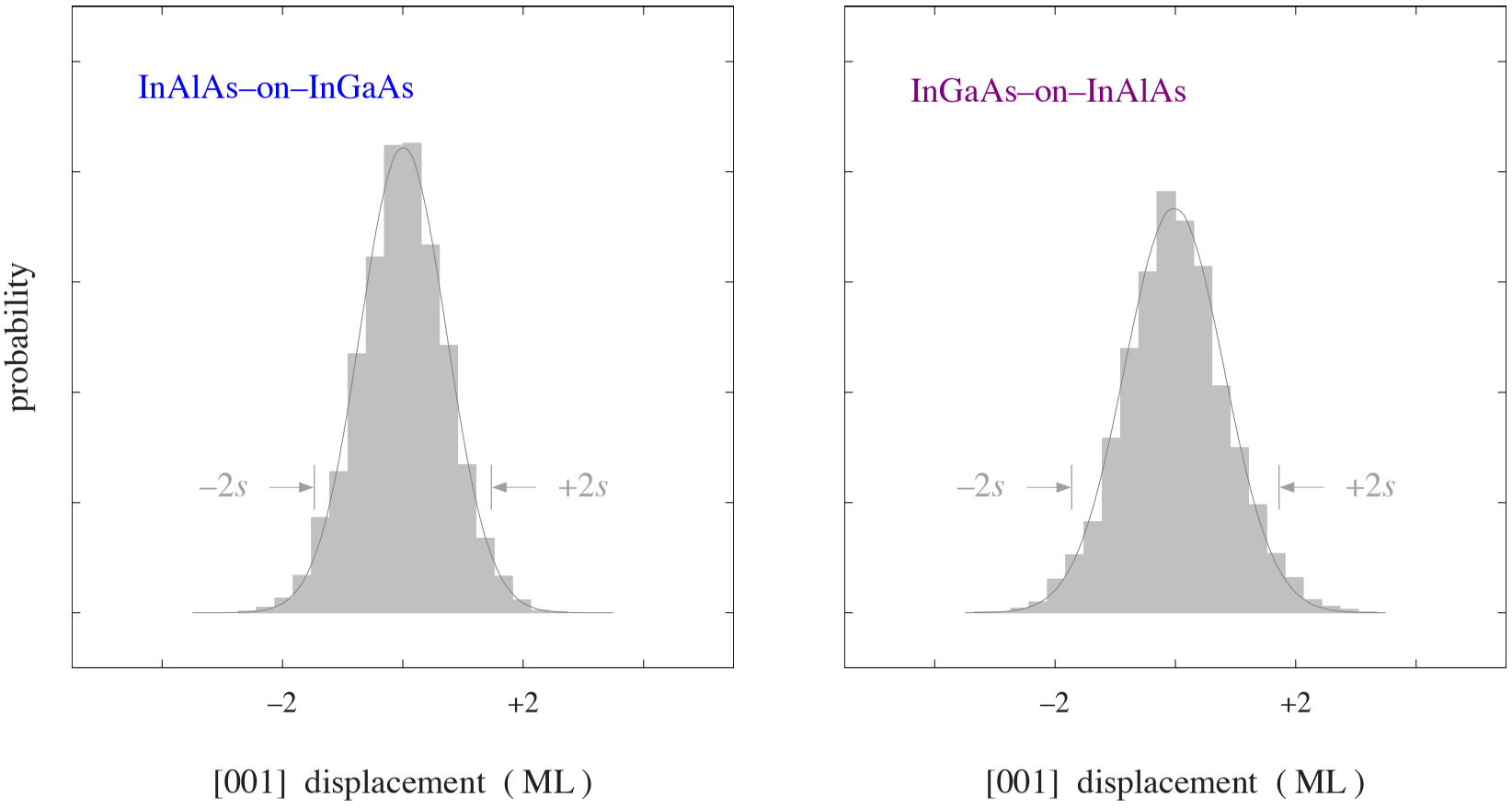


FIGURE 4.1. Distribution of vertical interface fluctuations, z_i, referenced to respective segment means and normalized to the local [001] lattice constant, from period–pooled, –2.50 V surveys in (110) cross section, together with best–fit Gaussian overlay (grey). Sample variance, s^2, is calculated via (4.6) and summarized in the table shown in page 178.

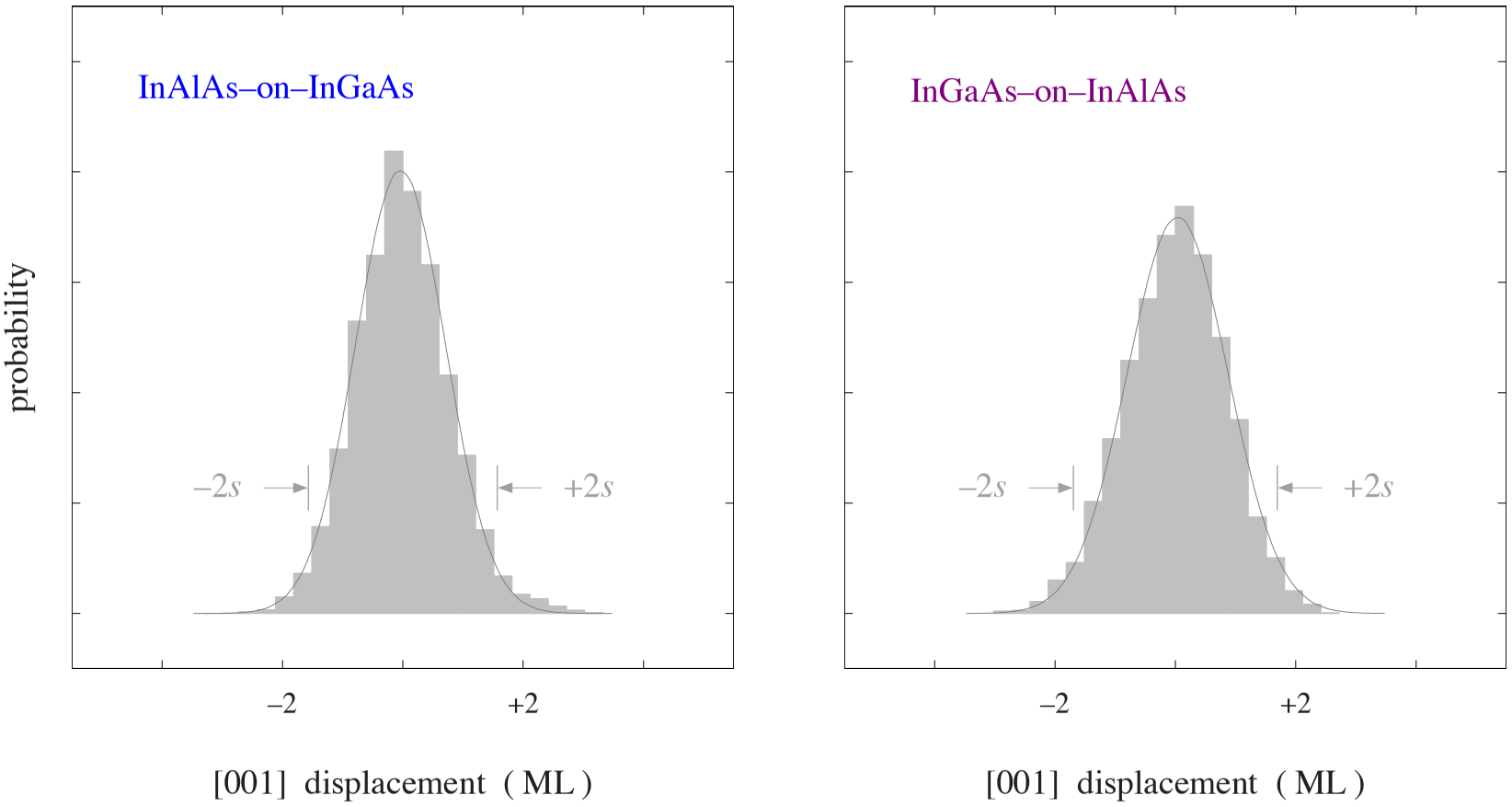


FIGURE 4.2. Distribution of vertical interface fluctuations, z_i, referenced to respective segment means and normalized to the local [001] lattice constant, from period–pooled, –2.50 V surveys in (1–10) cross section, together with best–fit Gaussian overlay (grey). Sample variance, s^2, is again calculated via (4.6) and summarized in the table shown in page 178.

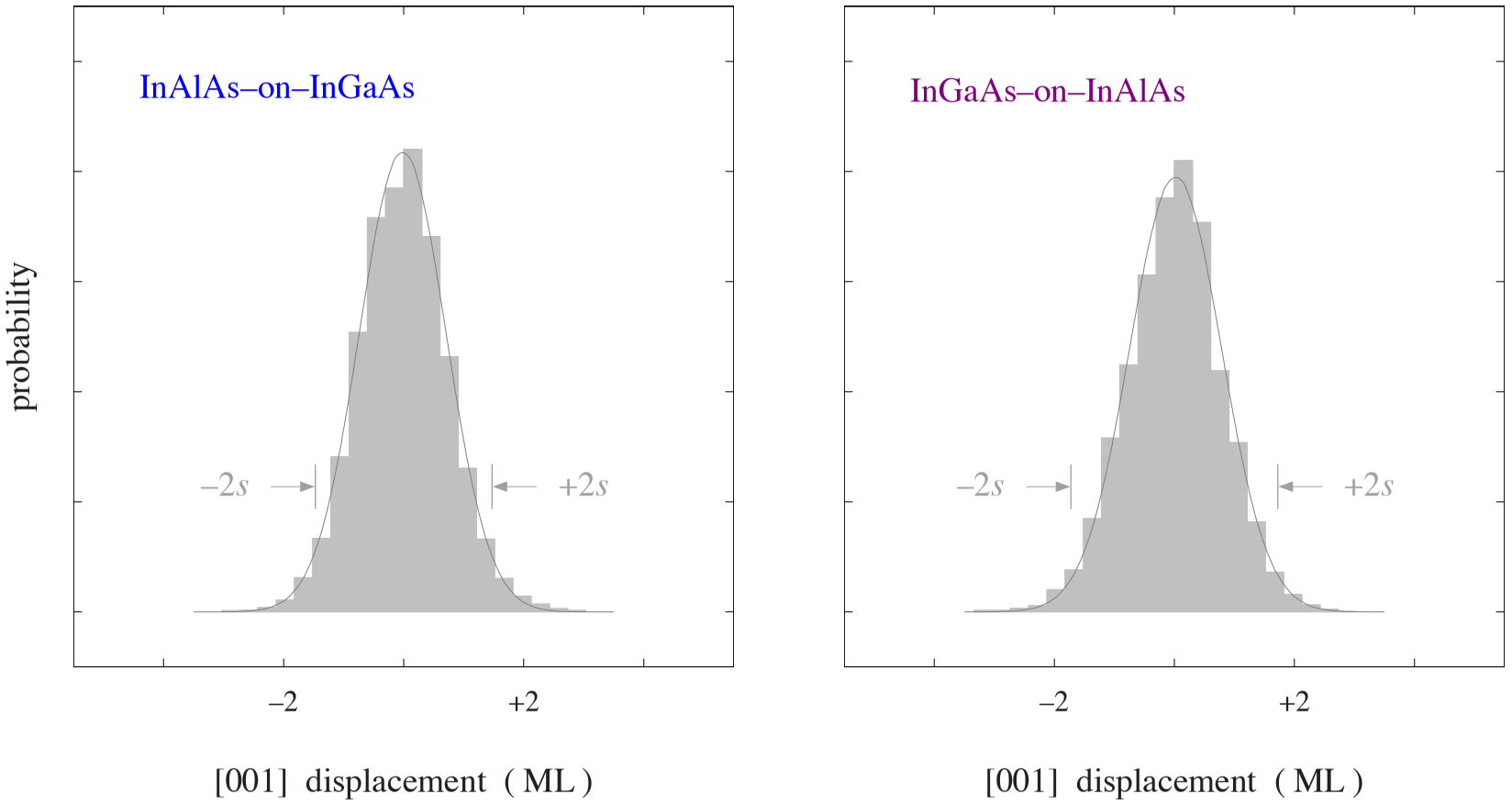


FIGURE 4.3. Distribution of vertical interface fluctuations, z_i, referenced to respective segment means and normalized to the local [001] lattice constant, from period–pooled, –2.25 V surveys in (110) cross section, together with best–fit Gaussian overlay (grey). Sample variance, s^2, is again calculated via (4.6) and summarized in the table shown in page 178.

and j are point (pixel) and segment indices for the specified ensemble, respectively – and each point, z_i, in the ensemble is referenced to the corresponding segment mean, $\langle z \rangle_j$, from which it is drawn[2].

Several important conclusions may be drawn from these histograms. First, as evidenced by the overlaid fits in each case, the fluctuations at either interface are Gaussian, a characteristic universally assumed but only directly confirmed with these plots. Second, as judged by the width of their respective distributions, InGaAs–on–InAlAs interfaces appear fundamentally different (i.e., rougher) from InAlAs–on–InGaAs ones[3]. This observation is significant insofar as any asymmetry in the statistical properties of the two heterojunctions translates into distinct scattering rates for the carriers entering barrier regions versus those exiting from them. Finally, both (110) and (1–10) cross sections are qualitatively similar, offering no evidence, here, for any strong, growth–plane anisotropy.

Simulated Profiles: Model Description

The naive approach to gauging roughness amplitudes from the corresponding widths of sampled displacement distributions suffices only when the stochastic variable in question is completely random and each measurement (i.e., realization) independent

[2] Although the interface profiles, themselves, are re–sampled to 400 ppi per the discussion in Chapter III, four–point bins – corresponding to the native one–angstrom pixel resolution with which the data was originally acquired – were employed in constructing the histograms displayed in Figs. 4.1 – 4.3.

[3] since these distributions are each normalized to unit area, their width and height are inversely related

of all others (i.e., no correlations in the data). As we now show, the inescapable referencing of segment fluctuations to individual segment means when sampling a correlated, non–deterministic process has far–reaching consequences for otherwise–trivial statistical quantities such as segment–mean fluctuations, variance of the segment means, variance of the fluctuations per segment, and variance of the segment variances. The latter two are of special importance in our studies since they directly bear on the experimental measurement of the roughness amplitudes and their uncertainties.

The unusual challenges associated with sampling a correlated stochastic process are best understood, and illustrated, by way of numerical simulation – that is, by applying the same protocols used for our analysis of experimental interface profiles (whose parent distributions are, by definition, unknown) to an (equivalent) ensemble of simulated profiles generated from parent distributions with *a priori* known statistical properties.

We restrict attention, for now, to profiles parameterized by a single, Gaussian correlation length; this is a common assumption in the literature [12 – 14] whose appropriateness is examined more carefully, later, in Chapter V. For reasons made clear below, it is simplest to begin this discussion in reciprocal space, where the presumed, parent power spectrum may be specified by way of a one–dimensional, analytic expression

$$\sqrt{2\pi}\ \Delta^2\ \Lambda\ \exp\left(-2\pi^2\ \Lambda^2\ q^2\right), \tag{4.1}$$

with Δ the roughness amplitude (Δ^2 the variance), Λ the Gaussian correlation length, q the spatial frequency ($|\kappa' - \kappa|$ being the corresponding momentum transfer, c.f. (1.7).

As written, the parameters appearing in (4.1) are understood to be physically–dimensioned quantities (e.g., units of Å or $Å^{-1}$), but for purposes of numerical simulation over a definite interval (in both reciprocal and direct space) they are re–interpreted as either integer quantities – or as dimensionless ratios – denominated in "STM" units of pixels. The discretized power spectrum for real–space profiles of integer length L is then, accordingly, given by

$$\sqrt{2\pi}\ \frac{\Delta^2}{d^2}\ \frac{\Lambda}{L\,d}\ \exp\left(-\frac{2\pi^2\ \Lambda^2\ q^2}{L^2\ d^2}\right), \tag{4.2}$$

where d is the physical pixel spacing, in units of Å, Δ as well as Λ are both re–scaled to d, and q is now integer.

Simulated profiles corresponding to particular, direct–space realizations of the so–digitized power spectrum are then generated from (4.2) as follows: with even integer L, we let q adopt all integer values between ($-L/2$, $L/2-1$); the dc–component at $q = 0$ is then removed from the digitized power spectrum – to ensure the corresponding real–space profiles have 0 mean – and the remaining ac power renormalized so that its area equals the desired, real–space variance, Δ^2 (now dimensionless). The (positive) square root of this renormalized power spectrum serves as the real part of a complex–valued simulation transform after assignment of a distinct, randomly–selected phase to each q–

point in the positive–half ($q > 0$) of reciprocal space; phases of identical magnitude, but opposite sign, are assigned each q of the negative–half ($q < 0$) to ensure our simulation transform is its own complex conjugate and all real–space profiles computed from it manifestly real. Each complex–valued spectrum so realized by sampling the available universe of random phases is then inverse–transformed to provide a particular, L–point representation from the mirroring ensemble of Gaussian–correlated, real–space profiles.

As previously mentioned, our experimental ensembles are assembled from disjoint, period–pooled, 350–point samplings from (roughly) micron–long surveys of each interface type, in specified cross section, with specified imaging voltage. To most nearly reflect similar circumstances in our numerical simulations, the number of points, L , was chosen to be 10,500, a value close to "one micron" but divisible into (an integer number of) 350–point segments[4]. This numerical ensemble, although of comparable size to our period–pooled experimental ones (see Table 4.1 later in this chapter), nevertheless proved insufficient for its intended purpose, and was later augmented with eight additional (disjoint, one–micron) simulations to provide the desired statistical precision. This enlarged, simulation ensemble represents two–to–three times the total experimental

[4] Though computationally more efficient to simulate an ensemble of individual, 350 Å segments instead of cutting up lengthier profiles to generate an equivalent number, this approach fails to properly mimic the effects of experimental sampling we seek to understand. This is due to the (cyclic) boundary conditions imposed by our computations on the ends of each segment; the influence of these convenient, but artificial, conditions "recedes" correspondingly as the simulated profile expands in length from one– to thirty–segments.

data available from pooling together the interface–type resolved surveys in Figs. 4.1 – 4.3 without regard to imaging voltage or cross section[5].

Simulated Profiles: Effect of Sampling on *a priori* Fluctuation Amplitude Distributions

To develop our intuition for how experimental sampling influences the roughness statistics exhibited by correlated data, we first examine the more familiar case of uncorrelated (i.e., randomly–distributed) profiles whose power spectrum is frequency–independent (i.e., white noise).

The variance (squared roughness amplitude) of any given, N–point segment, as well as the variance of the variance (squared uncertainty in the squared roughness amplitude) over any given, M–segment ensemble, is readily calculated from the corresponding set of $M \bullet N$–point simulated (real–space) profiles via

$$s_j^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} \left(z_i - \langle z \rangle_j \right)^2, \tag{4.3}$$

and

[5] Although the respective probability distributions portrayed in these figures appear remarkably similar, they cannot be legitimately pooled (at this point) for reasons we make clear in the following section. The data from each experimental survey in the table shown in page 178 is accordingly capped at ~1/10 our total simulation pool. As we will see, shortly, this is not as disappointing as first appears, since the central limit theorem reduces the discrepancy in signal–to–noise ratios to the square root of this number.

$$\mathrm{var}\left(s_j^2\right) \;=\; \frac{1}{M-1} \;\cdot\; \sum_{j=1}^{M}\left(s_j^2 - \left\langle s^2 \right\rangle\right)^2, \tag{4.4}$$

respectively. Here, z_i are the (dimensionless) vertical fluctuations of points *i*, belonging to segment *j*, about their respective segment mean,

$$\left\langle z \right\rangle_j \;=\; \frac{1}{N} \;\cdot\; \sum_{i=1}^{N} z_i, \tag{4.5}$$

and s^2 the mean (dimensionless) segment variance

$$\left\langle s^2 \right\rangle \;=\; \frac{1}{M} \;\cdot\; \sum_{j=1}^{M} s_j^2, \tag{4.6}$$

which serves as an estimate of the (dimensionless) parent variance, σ^2. To maintain clarity in the forgoing, as well as what follows, we have momentarily discarded the physical notation set out in Chapter I (and equation 4.1 here) in favor of a standardized notation prevalent throughout the statistics literature.

For the case of truly uncorrelated data, two, familiar and well–established expressions apply to the sampling statistics derived from an ideal ensemble ($M \rightarrow$ infinity) of *N*–point segments, namely

$$E\left[s_j^2\right] = \sigma^2, \tag{4.7}$$

and

$$E\left[\mathrm{var}\left(s_j^2\right)\right] = \frac{2\sigma^4}{N-1}, \tag{4.8}$$

respectively. The significance of these expectations is illustrated in Fig. 4.4 with a pair of 1400–point segments drawn at random from a much longer, uncorrelated profile simulated, in this case, with the aid of a random–number generator, assuming unit variance. The important lesson, here, is that the probability distribution of z–excursions referenced to the (respective) sample mean (black, blue) accurately represents the parent distribution (grey) in either case, so that the (Gaussian) widths of individual sample histograms correctly mirror the parent standard deviation within the uncertainty specified by (4.8).

A more instructive, and cautionary, example concerning the role of sampling is supplied by a parallel analysis of the profiles associated with correlated data. We begin, for concreteness, by examining the sampling statistics for 270 (*M*), 350–point (*N*) segments drawn from an ensemble of nine, 10,500–point simulated profiles, assuming a Gaussian–parent power spectrum with zero mean, unit variance, and 25–point correlation length, later extending our consideration to any correlation length within the allowable range $0 < \Lambda / N < 1 / 4$.

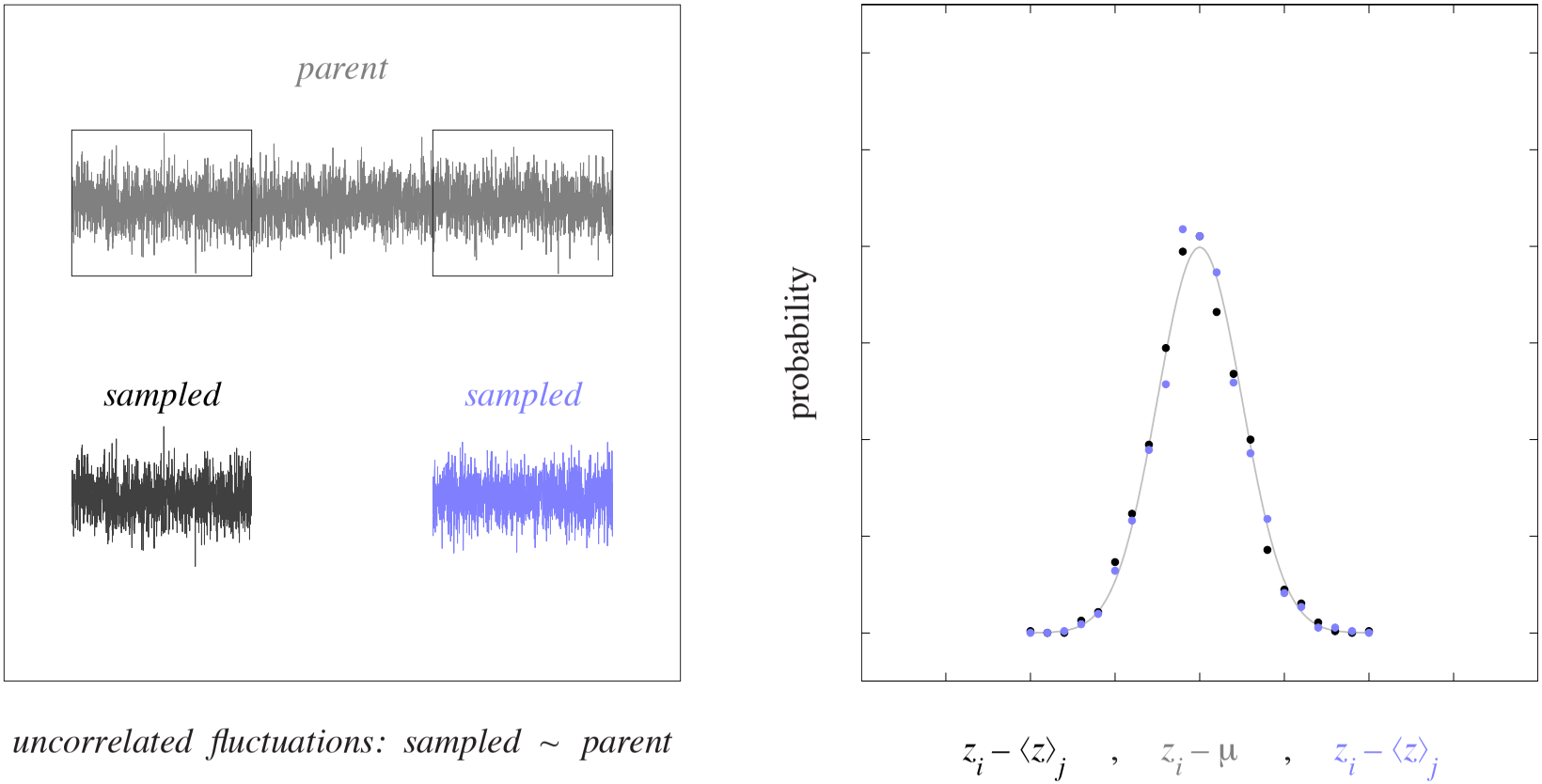


FIGURE 4.4. Simulated profile with random, uncorrelated fluctuations (left) and resulting distribution (right) of vertical displacements, z_i, relative to segment means for two, randomly–selected, sampling windows (black and blue). Sampled distributions accurately mirror the parent distribution (grey) in either case.

Fig. 4.5 presents the correlated–companion to Fig. 4.4, where two key differences emerge. First, the simulated, real–space profiles here (left) are notably bandwidth limited compared with those in Fig. 4.4; this is a general characteristic of correlated fluctuations and qualitatively reminiscent of our experimental data (cf., Fig. 3. 37, right). More importantly, as seen in Fig. 4.5 (right), the distribution of z–excursions deduced from any sampling window (black, blue) is unmistakably narrower than the (assumed) parent process (grey) would imply; thus, the finite–N segment variances for correlated data no longer, simply, mirror the population from which they are drawn.

To qualitatively appreciate why sampling might bias one's estimate of the true parent variance in this particular way, consider the difference between parent and individual–segment means indicated schematically in Fig. 4.6 (top) and (bottom). The parent variance is formally estimated via

$$\tilde{s}_j^2 \;=\; \frac{1}{N-1} \;\cdot\; \sum_{i=1}^{N} \left(z_i - \mu\right)^2 , \tag{4.9}$$

with

$$\left\langle \tilde{s}^2 \right\rangle \;=\; \frac{1}{M} \;\cdot\; \sum_{j=1}^{M} \tilde{s}_j^2 , \tag{4.10}$$

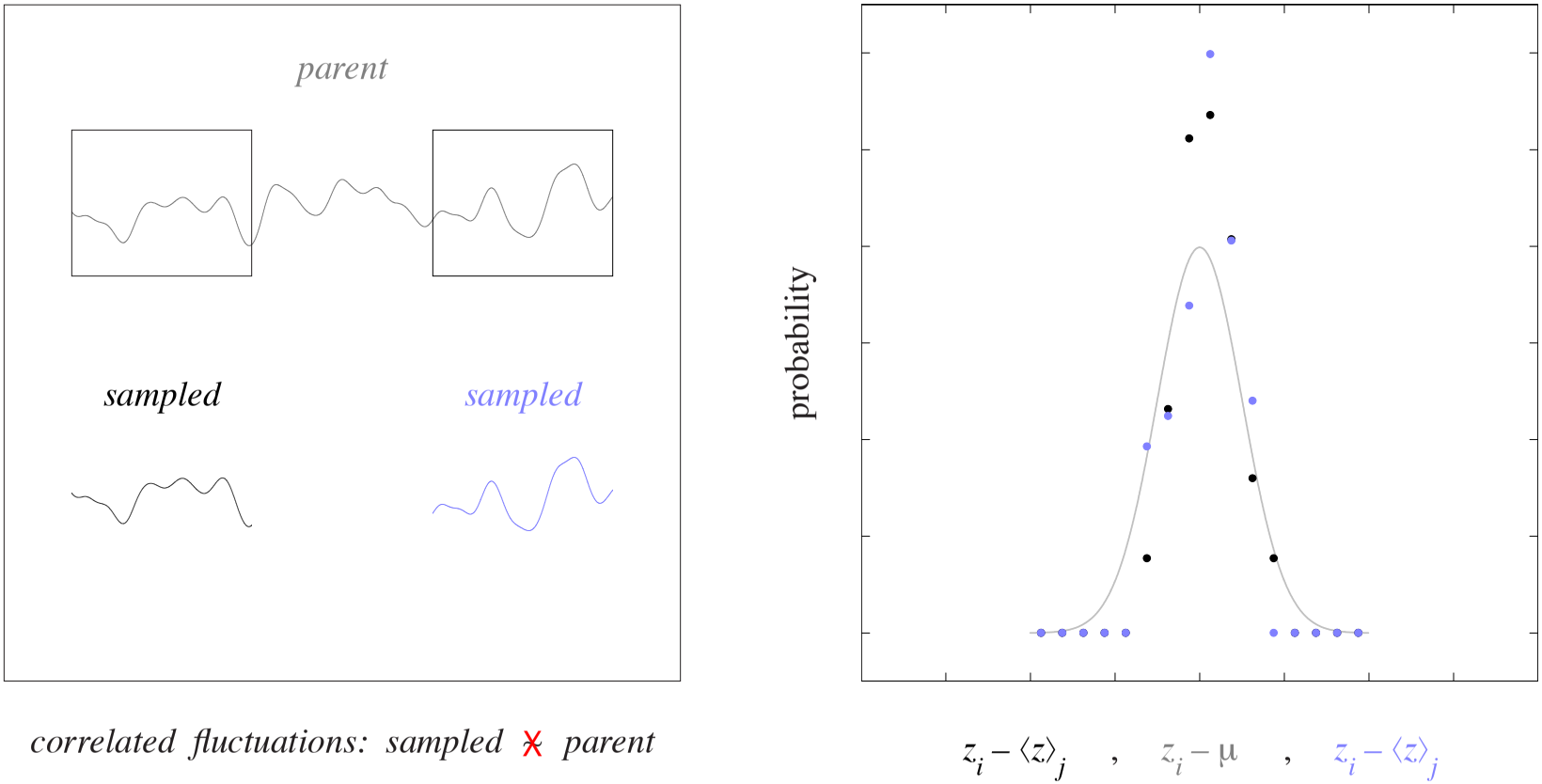


FIGURE 4.5. Same as Fig. 4.4, but for a Gaussian–correlated process. Narrowing of the sampled distributions relative to the (assumed) parent is evident.

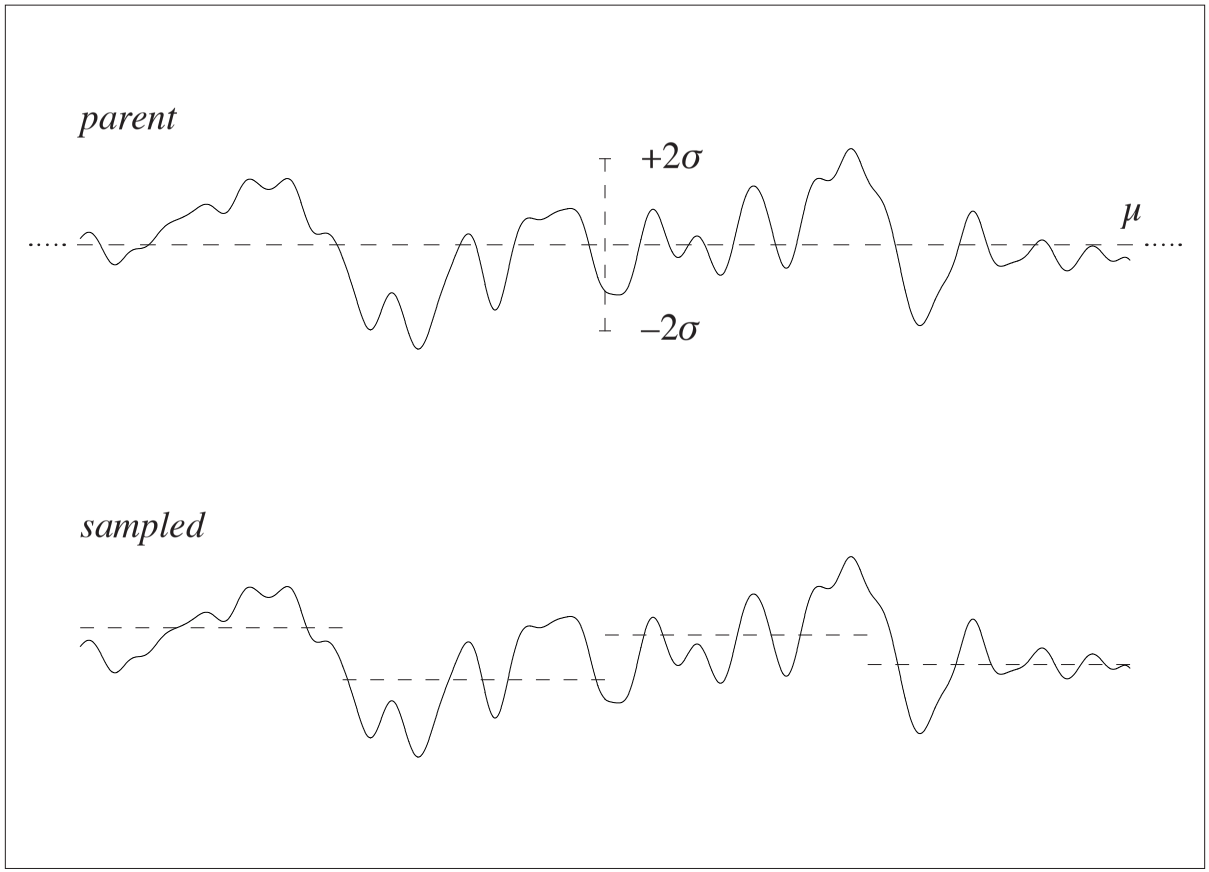


FIGURE 4.6. Simulated interface profile with Gaussian correlations (top) illustrating the distinction between a global, parent mean, μ, and individually–sampled, segment means (bottom) that serve as a local reference for vertical fluctuations in experimental interface profiles.

i.e, by referencing *z*–excursions to the (presumed known) parent mean, not individually–calculated segment means. A distinction arises between (4.9) and (4.3) when the individual segment means have failed to "effectively converge" to the parent mean; the (square root of the) sample variance, illustrated in Fig. 4.7 (bottom), is then reduced, relative to its parent value illustrated in Fig. 4.6 (top), by the fluctuation in this segment–to–segment reference. That intuition is explicitly corroborated in Fig. 4.8, where the simulated probability distribution (Λ / N = 25 / 350, M = 270) for $z_i - \mu$ is contrasted with the manifestly narrower distribution obtained for $z_i - \langle z \rangle_j$.

Two important questions naturally follow from this understanding: first, may a sample variance obtained from correlated data be appropriately "transformed" to reflect the (correct) underlying parent variance and, if so, how; second, what statistic(s) offers an appropriate and reasonable measure of the sampling uncertainty attached to this value.

Generalization of the numerical simulation just described to a uniformly–sampled (discrete) subset of correlation lengths ($0 < \Lambda / N < 1 / 4$) offers some insight into how the first of these questions might be addressed. Fig. 4.9 (left) presents the variance in the distribution of simulated segment means (discrete points) as a function of Λ / N, calculated via

$$\mathrm{var}\left(\langle z \rangle_j\right) \;=\; \frac{1}{M-1} \;\cdot\; \sum_{j=1}^{M} \left(\langle z \rangle_j - \langle z \rangle\right)^2 , \qquad (4.11)$$

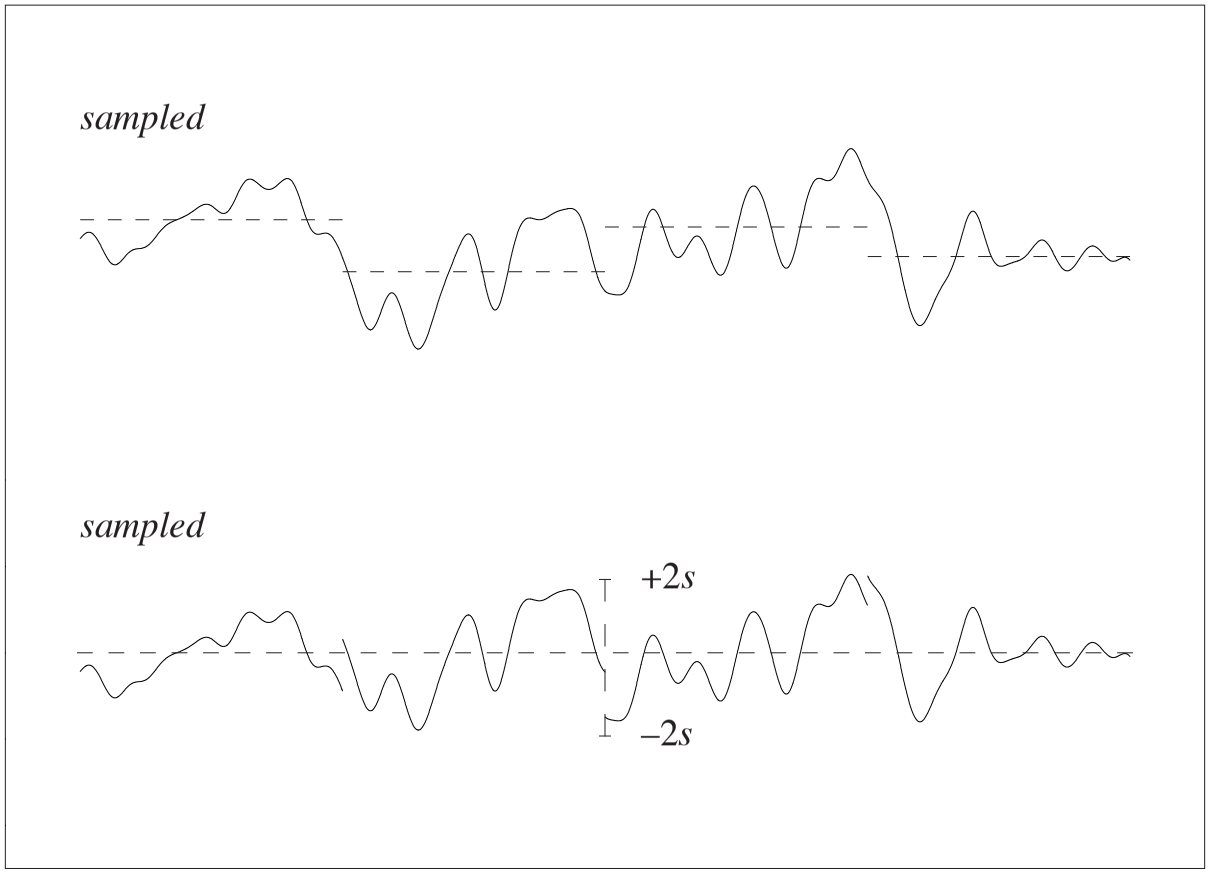


FIGURE 4.7. Continuation of Fig. 4.6, suggesting how a reduction in mean segment variance, *s*, relative to an assumed parent variance, σ, arises from reliance on individual segment means as vertical–fluctuation references.

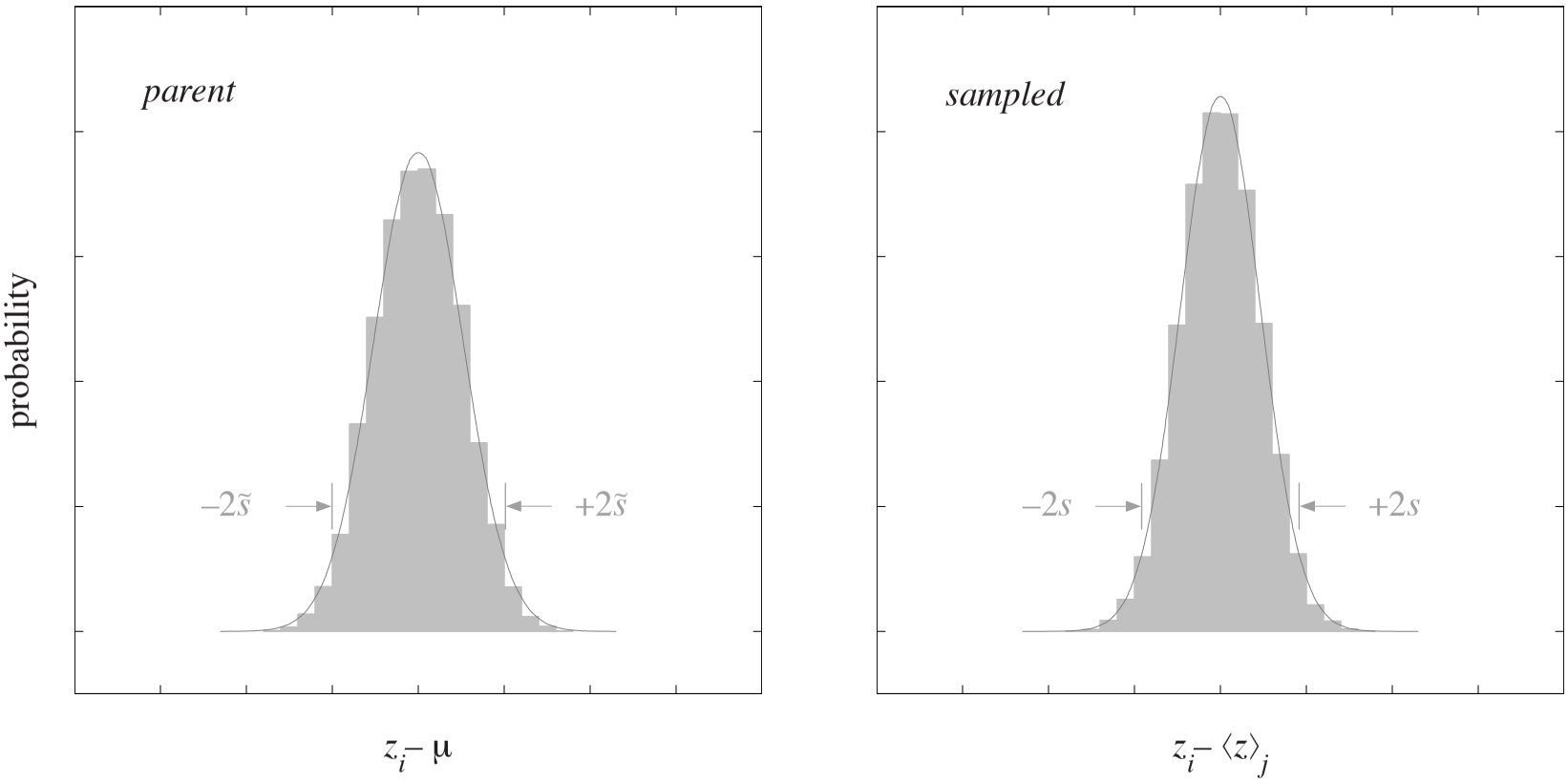


FIGURE 4.8. Probability distribution of vertical fluctuations, z_i, from an ensemble of simulated profiles with Gaussian correlations (Λ / N = 25 / 350), assuming these excursions are referenced either to the parent mean, μ, or individual segment means, $\langle z \rangle_j$. Corresponding sample variances are derived from (4.10) and (4.6), respectively. Grey curves are best–fit Gaussians.

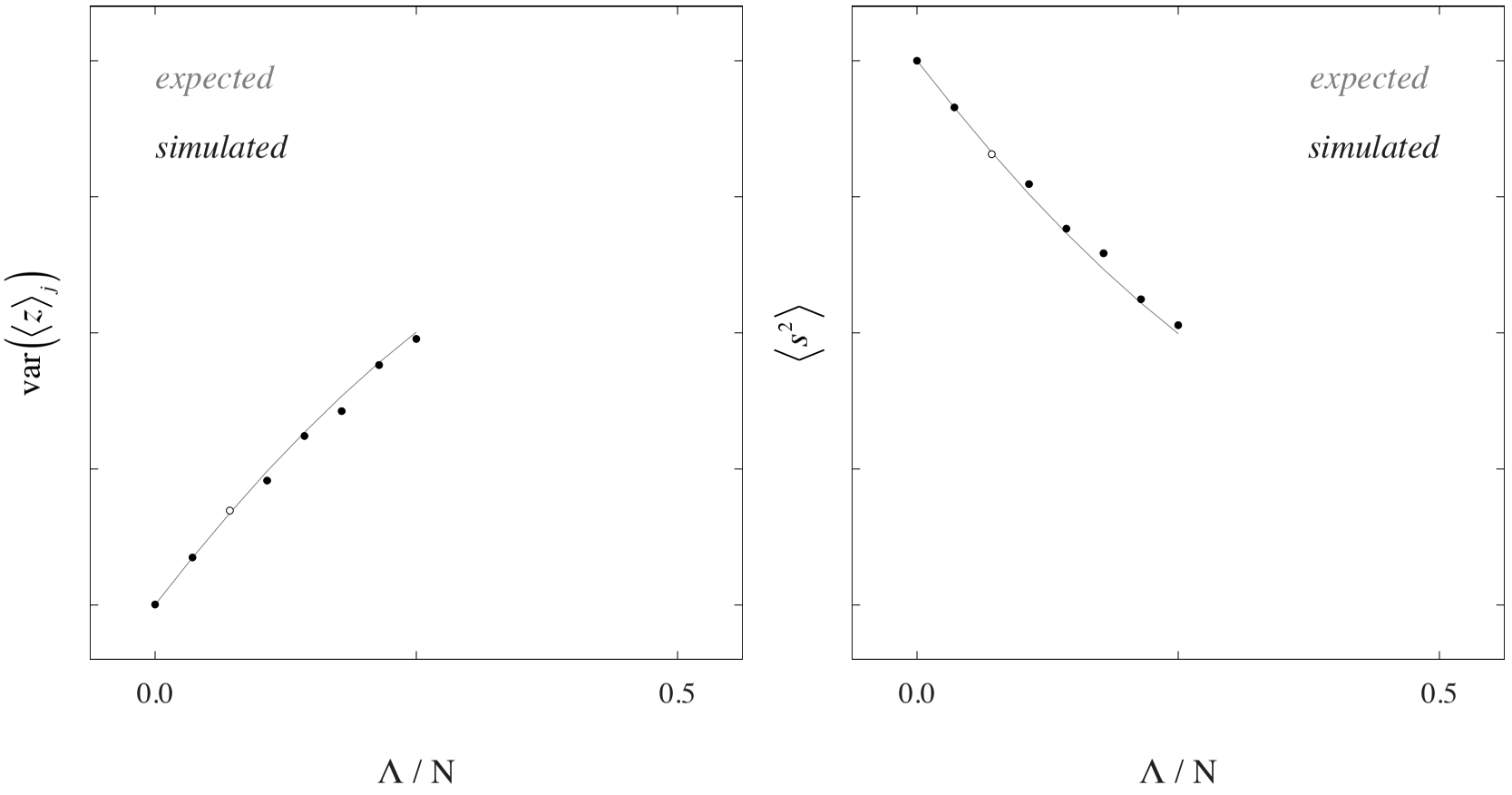


FIGURE 4.9. Variance of segment means, $\langle z \rangle_j$, and average segment variance, $\langle s^2 \rangle$, as functions of correlation length from numerical simulations (black), as well as expected behaviors (grey) predicted from (4.12) and (4.14) respectively. Vertical axes in both panels range from 0 to 1, reflecting normalization to unit parent variance. Open circle points to the specific case considered in Fig. 4.7.

where the (sampled) mean of all simulated segment means, $\langle z \rangle$, is equal to the parent mean (and set to zero) by construction. The expected value of this quantity, on the other hand, is given by a very general, theoretical expression[6],

$$E\left[\mathrm{var}\langle z\rangle_j\right] = \frac{\sigma^2}{N} \cdot \left[1 + 2\sum_{i=1}^{N-1}\left(1 - \frac{i}{N}\right)\rho_i\right], \tag{4.12}$$

valid for arbitrary (discrete) real–space, autocorrelation functions ρ_i. For Gaussian correlations, with

$$\rho_i = \exp\left(-\frac{i^2}{2\Lambda^2}\right), \tag{4.13}$$

this (theoretical) prediction (solid line in Fig. 4.9, left) agrees (as it should) with the corresponding M = 270 numerical simulation at each Λ / N following normalization to unit parent variance. Fig. 4.9 (right) presents a companion comparison of the mean sampled segment variance $\langle s^2 \rangle$, from (4.6), with its analogous theoretical expectation

[6] Priestley [58]M. B. Priestley, *Spectral Analysis and Time Series* (Academic Press, London, UK, 1981). offers a detailed explanation regarding the derivation of this equation, as well as (4.14) and (4.16).

$$E\left[s_j^2\right] \;=\; \sigma^2 \;\cdot\; \left[1-\frac{2}{N-1}\sum_{i=1}^{N-1}\left(1-\frac{i}{N}\right)\rho_i\right], \tag{4.14}$$

again following normalization to unit parent variance.

Fig. 4.9 (right) shows that as the correlation length Λ increases in relation to a fixed segment length N, the expected (as well as sampled) segment variance referenced to individual segment means decreases monotonically. Comparison with Fig. 4.9 (left) suggests this decrease is, at least qualitatively, mirrored by a corresponding increase in the expected (or sampled) variance in segment means, directly corroborating the intuition developed in Figs. 4.6 and 4.7. In particular, Fig. 4.9 (right) smoothly approaches unity (i.e., no distinction between sampled and parent variances) in the uncorrelated limit Λ / $N \rightarrow 0$, whereas Fig. 4.9 (left) smoothly approaches zero (i.e., vanishing dispersion of segment means about the parent mean) underscoring the fundamental distinctions between correlated and uncorrelated data.

A more interesting, and useful, insight concerning Fig. 4.9, however, is the following upper bound, established numerically upon direct substitution of (4.13) into (4.12) and (4.14),

$$E\left[s_j^2\right] \;+\; E\left[\mathrm{var}\langle z\rangle_j\right] \;\leq\; \sigma^2 \;\cdot\; \left(1+\frac{1}{N}\right), \tag{4.15}$$

where equality (rigorously) holds in the uncorrelated ($\rho_i \rightarrow 0$) limit; left and right panels

in Fig. 4.9 therefore sum to 1 with a precision better than $1/N$ for all nonzero Λ. One may therefore use (4.14) to confidently deduce a true parent variance from an infinite ensemble of sample variances (referenced to individual segment means) provided our assumed form of the correlation function (4.13) is correct and the corresponding value of Λ (abscissa in Fig. 4.9) known.

Having thus addressed the first of our two key questions, we turn attention, next, to the second. The distribution of segment variances – a histogram not ordinarily examined in the case of uncorrelated data – plays a central role here, since it provides an empirical estimate for the expected uncertainty attached to any single segment variance, s_j^2, calculated from (4.3) with correlated data. As it turns out, a single statistic from this histogram – the variance of segment variances – suffices for our purposes.

Fig. 4.10 (left) contrasts the Λ / N dependence of the expected variance in segment variances relative to the parent mean,

$$E\left[\mathrm{var}\left(\tilde{s}_j^2\right)\right] = \frac{2\left(\sigma^2\right)^2}{N} \cdot \left[1 + 2\sum_{i=1}^{N-1}\left(1 - \frac{i}{N}\right)\rho_i^2\right], \tag{4.16}$$

using (4.13) (solid line), with our M = 270 simulation results (discrete points), referenced to segment means using (4.4), following normalization to unit parent variance. (4.16) agrees with our segment–mean referenced simulations in the uncorrelated limit (Λ / $N \rightarrow 0$) but deviates ever more, as Λ / N increases, on account of

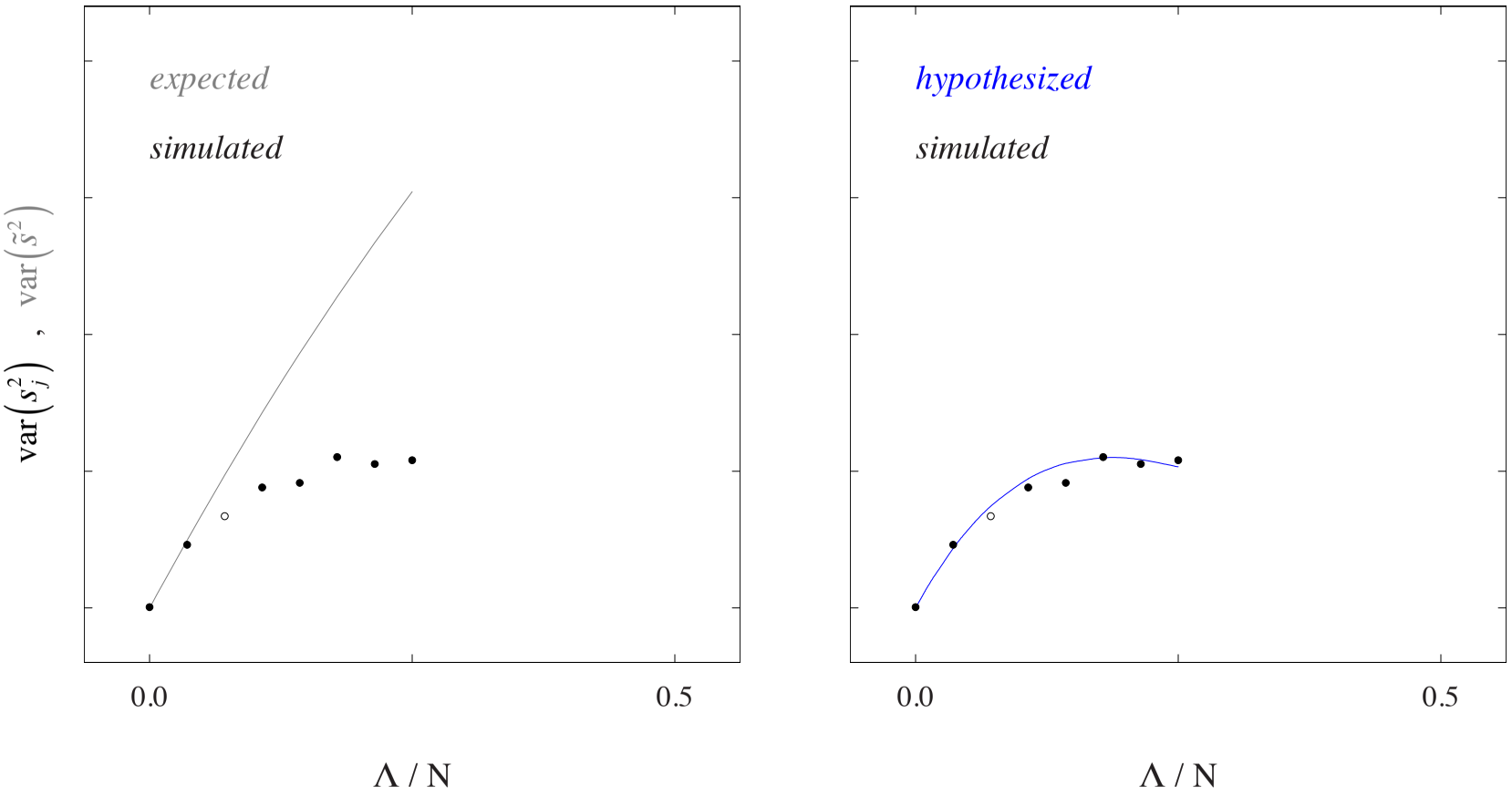


FIGURE 4.10. Variance in individual segment variances, s^2_j, as a function of correlation length, from numerical simulations referenced to local segment means (black), and (4.16) referenced to the global, parent mean (grey). Replacement of (4.16) by (4.18) in the right panel evidently accounts for this difference between global and local means. Vertical axes in both panels again range from 0 to 1.

the monotonically increasing dispersion in segment means illustrated in Fig. 4.9 (left). We are unaware of any analogous expression for the appropriate expectation for our purposes,

$$E\left[\mathrm{var}\left(s_j^2\right)\right], \tag{4.17}$$

but an intuition suggested by Fig. 4.9 (left) and numerically substantiated in Fig. 4.10 (right), namely that

$$E\left[\mathrm{var}\left(s_j^2\right)\right] \equiv E\left[\mathrm{var}\left(\tilde{s}_j^2\right)\right] - 2\left(E\left[\mathrm{var}\langle z\rangle_j\right]\right)^2, \tag{4.18}$$

appears reasonable. We presume this relation to be rigorously correct in the $M \rightarrow$ infinity limit, and henceforth adopt it as our working replacement for (4.16).

Irrespective of the validity of (4.18), we may then confidently conclude

$$\frac{\mathrm{var}^{1/2}\left(\left\langle s^2\right\rangle\right)}{\left\langle s^2\right\rangle} = \frac{1}{\sqrt{M}} \cdot \frac{\left(E\left[\mathrm{var}\left(s_j^2\right)\right]\right)^{1/2}}{E\left[s_j^2\right]}, \tag{4.19}$$

where the $1 / M^{1/2}$ reduction in relative sampling uncertainty (Law of Large Numbers) is guaranteed under no stronger assumption than all pooled segments, j, are identically distributed and likewise statistically independent of one another. It is important to note

the ratio of expectation values in (4.19) is a dimensionless function of Λ / N, and entirely independent of σ.

Figure 4.11 illustrates the predicted error bounds and expected segment variances as functions of Λ / N, from the right hand side of (4.19), for a simulation ensemble (M = 270), and a typical experimental ensemble (M = 36), assuming Gaussian correlations – (4.13) together with (4.18). The near–convergence of finite–ensemble simulations and infinite–ensemble expectations in Fig. 4.9 (right) appears consistent with the comparatively small dispersion seen here, in Fig. 4.11 (left), but the correspondingly–larger uncertainties anticipated with smaller, experimental ensembles (Fig. 4.11, right), present a more sobering picture.

Practical implementation of these concepts with finite–M experimental ensembles requires replacement of the expectation values appearing on the right–hand–side of (4.19) with their corresponding M–segment estimates via

$$\frac{\mathrm{var}^{1/2}\left(\left\langle s^2 \right\rangle\right)}{\left\langle s^2 \right\rangle} = \frac{1}{\sqrt{M}} \cdot \frac{\mathrm{var}^{1/2}\left(s_j^2\right)}{\left\langle s^2 \right\rangle}, \qquad (4.20)$$

Experimental Profiles: Summary of Observations

Having thoroughly addressed the statistical subtleties associated with the sampling of correlated stochastic processes, we are now well positioned to summarize the mean sample variances deduced from our experimental ensembles (Figs. 4.1 – 4.3) and assess their significance in light of appropriately–assigned experimental

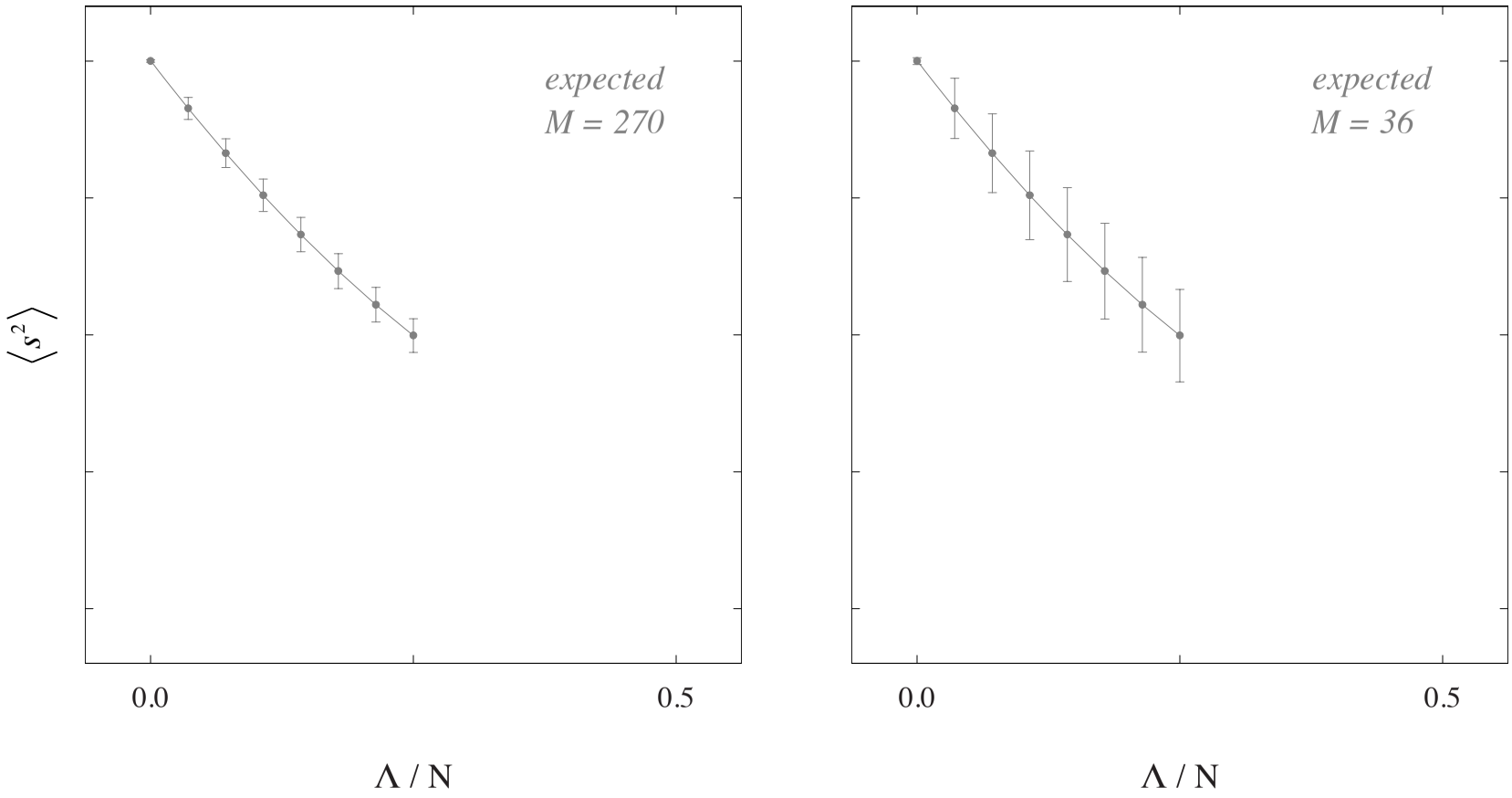


FIGURE 4.11. Mean segment variance versus correlation length, with error bounds from (4.18), for representative simulation ($M = 270$) and experimental ($M = 36$) ensembles. Vertical axes in both panels range from 0 to 1, reflecting normalization to unit parent variance.

uncertainties.

The results are summarized in Table 4.1. Two, general, comments are in order. First, the relative uncertainties computed via (4.20) exceed those one would otherwise assign uncorrelated data by an order of magnitude, and also exceed those found from least–squares Gaussian fits (Fig. 4.1 – 4.3) by (roughly) a factor of four. Second, these relative uncertainties, being comparatively small (between 5.0 and 7.5 %), bracket the second point ($\Lambda / N = 12.5$[7]$/ 350$) in Fig. 4.11 (right). This unexpectedly "short" estimate for the correlation length – assuming the Gaussian dependence given in (4.13) – fits comfortably within our sampling window, a surmise likely true irrespective of this precise functional form in view of (4.20)'s dependence solely on Λ / N. It likewise suggests our observed sample variances may only differ by 10 % (or so) from their true parent variances. As already emphasized above, the relative uncertainties in Table 4.1 are independent of this (or any) estimate of parent values.

That our data imply $\Lambda / N \leq 0.1$ demonstrates the self–consistency of our previous assumption segments taken from successive images – whose separation is of order $2N$ – are statistically independent of one another, but offers no similar assurance period–pooled segments from within the same image behave likewise.

The following conclusions with respect to the statistical significance of differences in segment variances between various pairs of ensembles (t–Test) in Table 4.1 hold under the assumption all M samplings are, in fact, independent of one another:

[7] 12.5 nominal Å correspond to approximately 3.3 [110] lattice constants.

Table 4.1. Mean segment variance, and associated error from (4.20), for period–pooled surveys at –2.50 V (black) and –2.25 V (grey) respectively.

	InAlAs–on–InGaAs		InGaAs–on–InAlAs	
	$\langle s^2 \rangle$ (ML^2)	segments (M)	$\langle s^2 \rangle$ (ML^2)	segments (M)
(110)	**0.53 ± 0.03**	40	**0.73 ± 0.04**	43
$(1\bar{1}0)$	**0.61 ± 0.05**	36	**0.71 ± 0.05**	34
(110)	**0.55 ± 0.03**	35	**0.61 ± 0.03**	46

(1) The existence of a growth–order–dependent asymmetry, distinguishing InGaAs–on–InAlAs heterojunctions (or so–called Al templates) from InAlAs–on–InGaAs ones (Ga templates), is conclusively established by the –2.50 V data in (110) cross section (~99.997 % confidence), very likely demonstrated by the –2.50 V data in (1–10) cross section (~94.21% confidence), and further substantiated by the –2.25 V data in (110) cross section (~92.94% confidence). In all cases, the mean segment variance (roughness) observed for Al templates measurably exceeds that of Ga templates.

(2) The mean segment variance of Ga templates in (110) cross section appears to be energy–independent over the narrow (0.25 eV), experimentally–accessible window; the mean segment variance of Al templates, on the other hand, is clearly (~99.07% confidence) energy–dependent over this same range, signaling an electronic contribution to the STM contrast defining the InGaAs–on–InAlAs heterojunction profile.

(3) There is no evidence to support a hypothesized crystalline anisotropy of Al templates, as the observed segment variances in (110) and (1–10) cross sections are statistically indistinguishable. On the other hand, there appears to be some, but not necessarily compelling, evidence (86.69% confidence) for a greater variance of Ga templates in (1–10) versus (110) cross section[8].

[8] An admittedly conceivable, though somewhat contrived, alternative interpretation – in view of the fact (110) and (1–10) data come from different (but close by) dies (Fig. 2.3) – is the roughness of Ga templates varies across the wafer whereas that of Al templates does not; this seems less likely than anisotropy, given the acknowledged temperature and flux uniformities associated with MOCVD versus MBE.

CHAPTER V

ROUGHNESS CORRELATION CHARACTERIZATION

Introduction

Thus far we have investigated only the roughness variances associated with our physically–motivated interface profile ensembles. Here we address the complementary dimension needed for a complete statistical description, namely the in–plane, spatial–frequency distribution of these [001] fluctuations, or the power spectrum. Our mathematical tool for generating power spectral densities (PSDs) from real–space interface profiles will be the cyclic, discrete Fourier transform. All of the PSDs described in this chapter are presented in (conventional) log power versus reduced wave vector format and normalized to unit area so their Fourier transforms yield real–space autocorrelation functions with unit variance: multiplication by an appropriate sample variance (Table 4.1) provides the corresponding real–space autocovariance.

Experimental Roughness Power Spectral Densities

We begin with the ensemble–averaged power spectral density obtained from Fourier transformation of individual InAlAs–on–InGaAs interface profiles in (110) cross section, and –2.50V sample bias, shown in Fig. 5.1 (left); this spectrum, along with those that follow, is plotted over the physically–relevant interval $0 \leq q \leq 1$, where q denotes

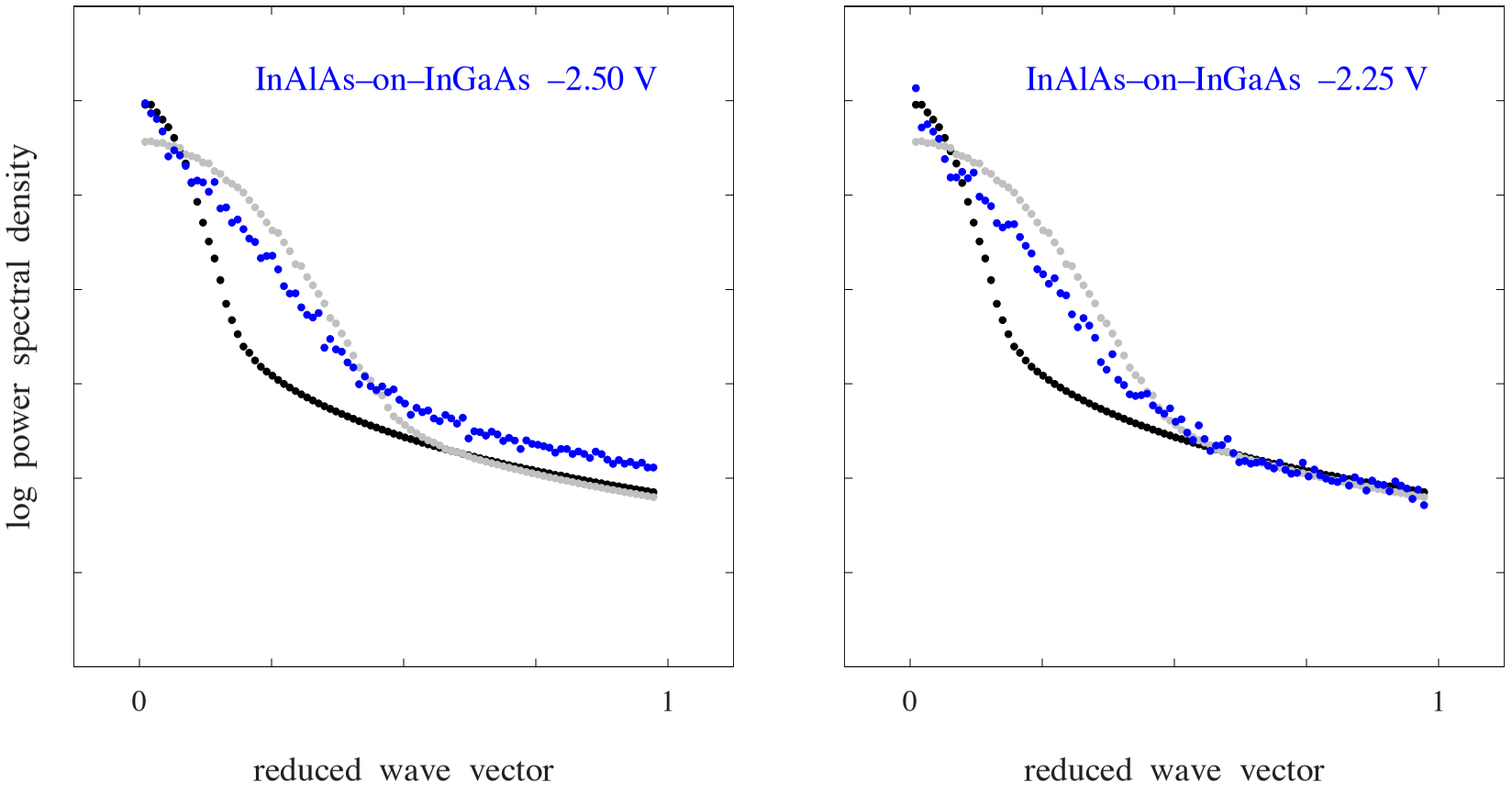


FIGURE 5.1. Ensemble power spectral densities (normalized to unit area) for the InAlAs–on–InGaAs heterojunction (blue) in (110) cross section. Like–normalized power spectra from simulated profiles with Gaussian correlation lengths of 3.3 (black) and 1.3 (grey) lattice constants are overlaid on both –2.50 V and –2.25 V data. The first correlation length is suggested by the relative uncertainty of the mean sample variance in Table 4.1, whereas the second conveniently brackets our data.

the magnitude of the appropriate ⟨110⟩–like wave vector. The mirroring PSD from our survey over the same area at –2.25V is shown in Fig. 5.1 (right).

Much of the discussion in Chapter IV – as well as in Chapter I – was built on the generally–accepted premise [12] that the real–space correlation functions describing our interface ensembles, along with their Fourier transforms, are Gaussian in nature. Fig. 5.1 explicitly demonstrates this is not the case. By comparing our InAlAs–on–InGaAs power spectra with the power spectra from identically–sampled, simulated profiles of bracketing Gaussian correlation lengths[1], we see no single, Gaussian process adequately describes these experimental data; the initial, linear fall–off at small wave vectors ($q \lesssim 0.5$) surprisingly indicates exponential behavior instead. In this connection, it is worth noting the near convergence of simulated and experimental power spectra to a seemingly "universal" functional form for $q \geq 0.5$ arises from finite–length sampling together with digitization, which introduce "brown" (with an inverse–square frequency fall off) and "white" (frequency–independent) noise, respectively, into an otherwise rapidly–decaying power spectrum. This convergence to a common shape simply re–affirms that all data, whether experimental or numerical, has been treated equivalently.

Focusing attention, then, on the physically–significant, small–wave–vector regime (save the vanishing $q = 0$ dc component), we see in Fig. 5.2 that exponential fits over the restricted range $0 < q \lesssim 0.30$ offer a persuasive match to these data. Corresponding analysis of the ensemble–averaged power spectral densities from

[1] With values of 3.3 (black) and 1.3 (grey) in–plane [110] lattice constants, respectively; the former value was suggested by the relative uncertainty of the mean sample variance given in Table 4.1.

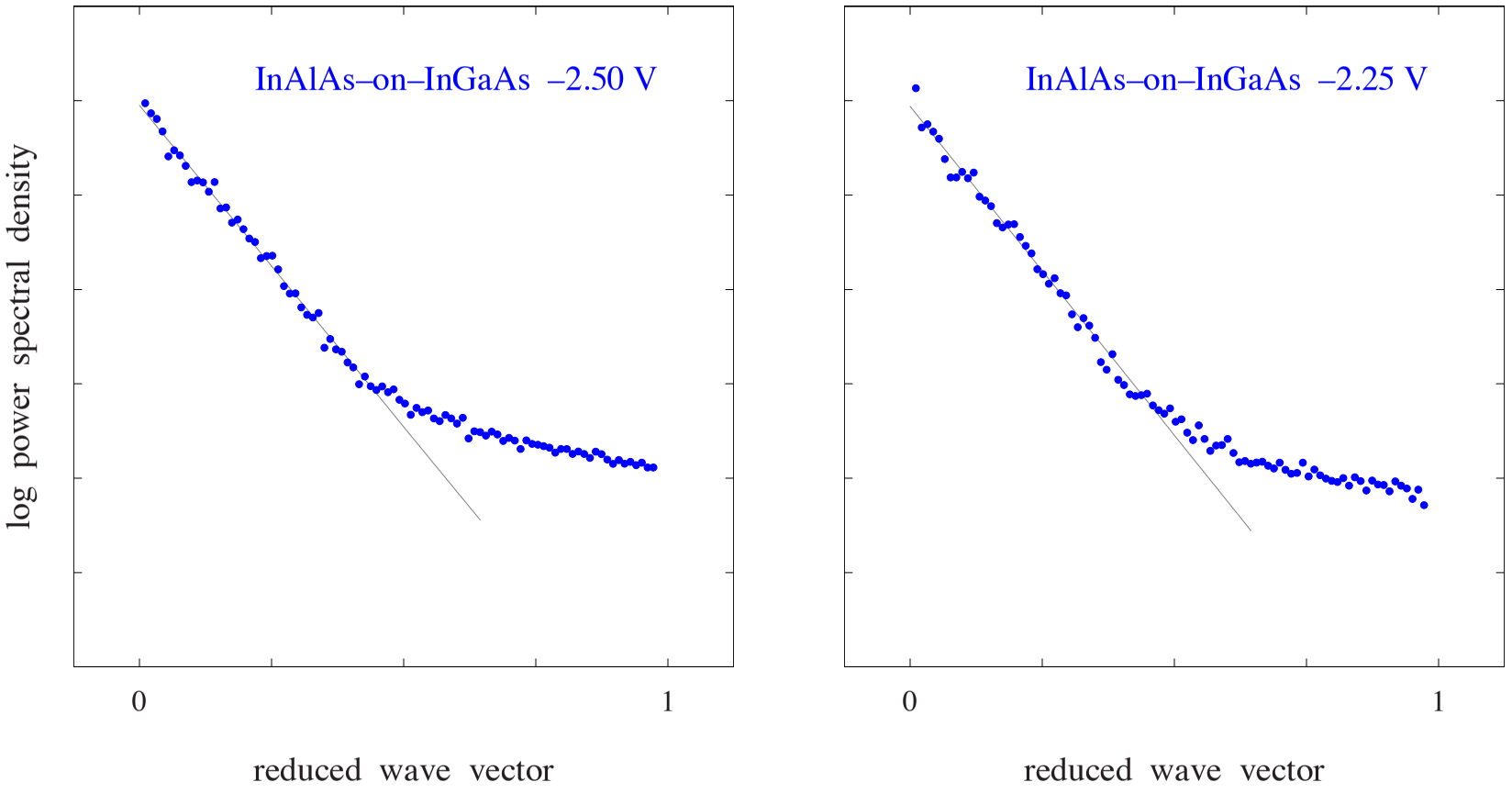


FIGURE 5.2. Power spectral densities as in Fig. 5.1, fit to a decaying exponential (grey) over the restricted (spatial) frequency range $0 < q < 0.30$. The excess power for $q > 0.5$ arises from sampling and digitization effects unrelated to physical properties of the heterojunction.

individual InGaAs–on–InAlAs interface profiles is summarized in Fig. 5.3. The two imaging voltages are, again, practically indistinguishable from one another, but the power spectra, here, appear qualitatively different from their InAlAs–on–InGaAs counterparts in Fig. 5.2; an overall exponential behavior still prevails, but one notes the emergence of a distinctly–steeper fall off in the power spectrum at very smaller wave vectors ($q \lesssim 0.05$), together with a discernible inflection in the curvature of the data at intermediate wave vectors ($0.05 \lesssim q \lesssim 0.30$) that argues for a possible (low–amplitude) Gaussian component as well.

The analogous power spectral densities in (1–10) cross section are summarized in Fig. 5.4. Here we note the seemingly different, growth–order–dependent behaviors exhibited in (110) cross section are reversed: the InAlAs–on–InGaAs interface now exhibits multi–component behavior whereas the InGaAs–on–InAlAs interface is accurately described by a single exponential. This puzzling result, which has no apparent crystallographic basis, effectively undermines any argument for growth–order–dependent power–spectral densities naively prompted by comparison of Figs. 5.2 and 5.3.

Setting this surprise aside for the moment, we turn, next, to some important mathematical details needed to connect the exponential power spectral density fits in Figs 5.2 – 5.4 with their associated, real–space correlation lengths. We begin with the comparatively straightforward example offered by a continuous, one–dimensional, Gaussian power spectral density whose real–space autocorrelation function (ACF) is another Gaussian, related to the first via (c.f. 4.1)

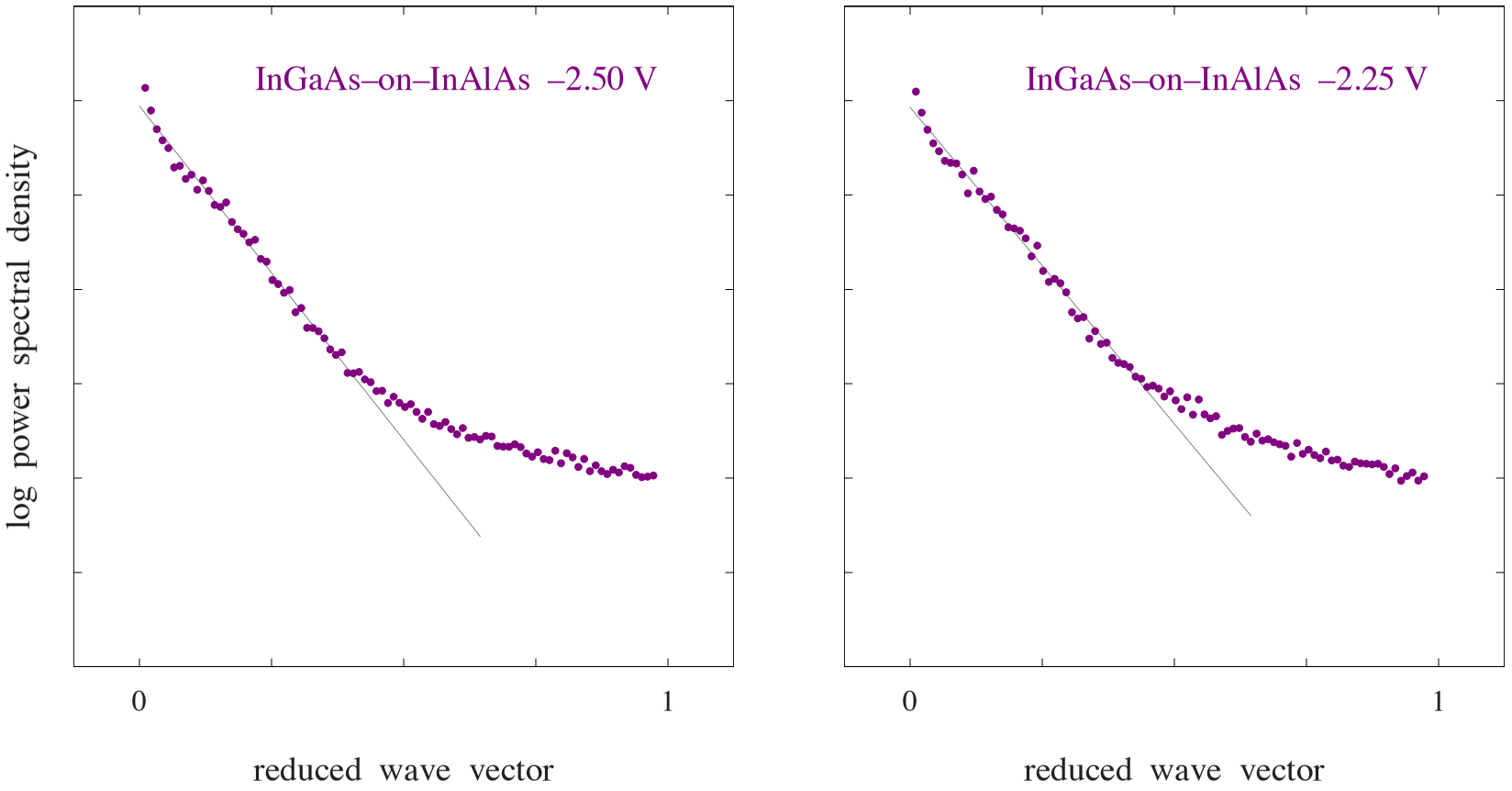


FIGURE 5.3. Ensemble power spectral densities (normalized to unit area) for the InGaAs–on–InAlAs heterojunction (plum) in (110) cross section, fit to decaying exponentials (grey) as in Fig. 5.2.

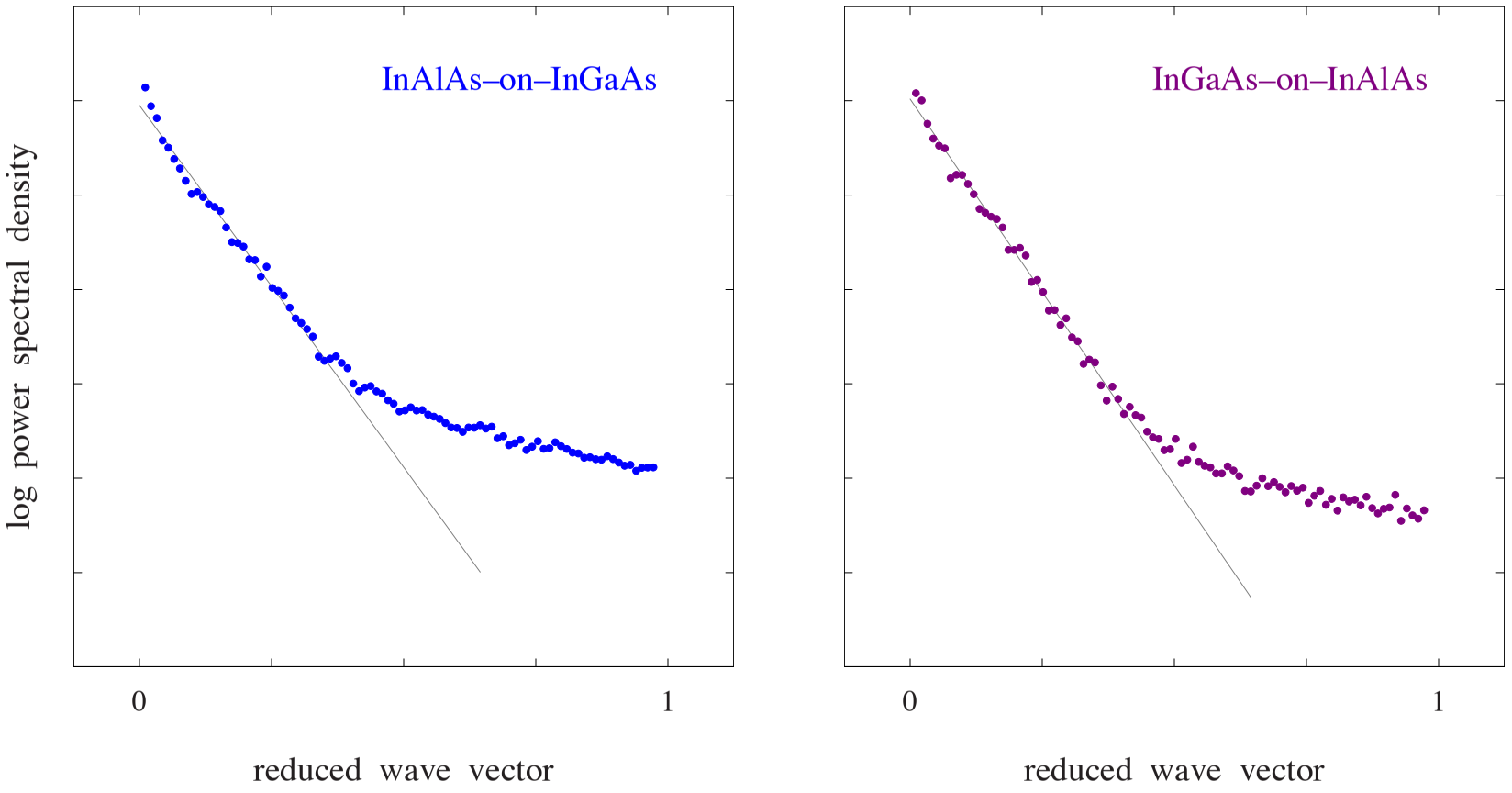


FIGURE 5.4. Ensemble power spectral densities (normalized to unit area) for the InAlAs–on–InGaAs (blue) and InGaAs–on–InAlAs (plum) heterojunctions in (1–10) cross section, fit to decaying exponentials (grey) as in Fig. 5.2.

ACF PSD

$$\exp\left(-\frac{x^2}{2\,\Lambda_x^2}\right) \quad \leftarrow \text{FT} \rightarrow \quad \sqrt{2\pi}\;\Lambda_x\;\exp\left(-2\pi^2\;\Lambda_x^2\;q^2\right), \tag{5.1}$$

where we use Λ_x to denote a real–space correlation length rather than a vector component. Defining an analogous, reciprocal–space correlation length Λ_k via

$$\Lambda_x = \frac{1}{2\pi\;\Lambda_k}, \tag{5.2}$$

we see the right–hand side of (5.1) may be re–written as

$$\frac{1}{\sqrt{2\pi}\;\Lambda_k}\;\exp\left(-\frac{q^2}{2\,\Lambda_k^2}\right), \tag{5.3}$$

thereby establishing (5.2) as the appropriate connection between real– (Λ_x) and reciprocal–space (Λ_k) quantities.

Like treatment of the discretely–sampled version of (5.1),

ACF PSD

$$\exp\left(-\frac{x^2}{2\,\Lambda_x^2}\right) \quad \leftarrow \text{FT} \rightarrow \quad \sqrt{2\pi}\;\frac{\Lambda_x}{L\;d}\;\exp\left(-\frac{2\pi^2\;\Lambda_x^2\;q^2}{L^2\;d^2}\right), \tag{5.4}$$

with L and d as previously defined in Chapter IV, then requires

$$\Lambda_x = \frac{L\ d}{2\pi\ \Lambda_k}, \tag{5.5}$$

for the right–hand side of (5.4) to again reduce to

$$\frac{1}{\sqrt{2\pi}\ \Lambda_k}\ \exp\left(-\frac{q^2}{2\ \Lambda_k^2}\right). \tag{5.6}$$

That (5.5) indeed provides the appropriate connection between real– and reciprocal–space correlation lengths with discretely–sampled, Gaussian power spectral densities is independently confirmed by numerical simulation.

Since the spectral density characterizing the (discretized) experimental profiles in Figs. 5.2 –5.4 appears predominantly exponential, it is natural to consider, next, a continuous, one–dimensional, exponential power spectrum whose Fourier transform partner is a Lorentzian autocorrelation function according to

$$\begin{array}{ccc} \text{ACF} & & \text{PSD} \\ \dfrac{1}{1+\dfrac{x^2}{\Lambda_x^2}} & \leftarrow \text{FT} \rightarrow & \pi\ \Lambda_x\ \exp\left(-2\pi\ |q|\ \Lambda_x\right), \end{array} \tag{5.7}$$

As before, replacing

$$\Lambda_x = \frac{1}{2\pi\ \Lambda_k}, \tag{5.8}$$

permits the right–hand side of (5.7) to be re–written as

$$\frac{1}{2\ \Lambda_k}\ \exp\left(-\frac{|q|}{\Lambda_k}\right). \tag{5.9}$$

Relying on the established relation between (5.4) and (5.1), we presume the discretely–sampled version of (5.7) to be

$$\begin{array}{ccc} \text{ACF} & & \text{PSD} \\ \frac{1}{1+\frac{x^2}{\Lambda_x^2}} & \leftarrow \text{FT} \rightarrow & \pi\ \frac{\Lambda_x}{L\ d}\ \exp\left(-\frac{2\pi\ |q|\ \Lambda_x}{L\ d}\right), \end{array} \tag{5.10}$$

and, likewise relying on (5.5), the discrete analog of (5.8) to be

$$\Lambda_x = \frac{L\ d}{2\pi\ \Lambda_k}, \tag{5.11}$$

so the right–hand side of (5.10) again reduces to

$$\frac{1}{2\,\Lambda_k}\,\exp\left(-\frac{|q|}{\Lambda_k}\right). \tag{5.12}$$

Although (5.10) – (5.12) are clearly motivated and self consistent, they have not been independently confirmed with numerical simulations in the same way as (5.4) – (5.6). We nevertheless use (5.11) to determine the real–space correlation lengths (Λ_x) corresponding to the exponential decay constants (Λ_k) obtained from our fits to the power spectra in Figs. 5.2 – 5.4, and these results are summarized in Table 5.1.

We note the real–space correlation lengths assembled in Table 5.1 are in surprisingly good agreement with the value of 3.3 lattice constants (~12.5 Å) previously estimated from the relative uncertainty in mean sample variance under the (now discredited) assumption of Gaussian statistics (Chapter IV) and illustrated in Fig. 5.1. It is likewise important to point out the quoted errors in Table 5.1 correspond to least–squares uncertainties and, based on our exhaustive analysis of roughness histograms in Chapter IV, these could easily underestimate the "true" statistical errors by as much as a factor of four. We do not yet have a framework for assessing the influence of sampling statistics on power spectral density estimation comparable to our rigorous understanding of sample variance. For this reason, as well as additional considerations that follow below, we abstain, here, from any statistical test of the significance of differences between mean (roughness) correlation lengths analogous to that pursued with regard to mean (roughness) variances in Table 4.1.

Table 5.1. Correlation lengths (Λ) obtained from exponential fits to the period–pooled, power spectra in Figs. 5.2 – 5.4 for surveys at –2.50 V (black) and –2.25 V (grey) respectively. Quoted uncertainties are uncorrelated, least–squares fitting errors and therefore likely too small.

	InAlAs–on–InGaAs		InGaAs–on–InAlAs	
	Λ ($\mathbf{a}_{110}$)	segments (M)	Λ ($\mathbf{a}_{110}$)	segments (M)
(110)	**2.50 ± 0.06**	40	**2.59 ± 0.08**	43
$(1\bar{1}0)$	**2.82 ± 0.08**	36	**3.02 ± 0.07**	34
(110)	**2.56 ± 0.08**	35	**2.46 ± 0.07**	46

We now return to the challenge posed by the relationship of the PSDs in Fig. 5.4 to those presented earlier in Figs. 5.2 and 5.3. Our statistical ensembles thus far have been motivated by physical factors generally acknowledged to play an important part in shaping interface properties, and that approach is re–affirmed with the ensemble–dependent variances, tabulated in Table 4.1, whose statistical significance further underscores these classifications. In short, the differences highlighted in Table 4.1 are undeniably real and of physical origin. But in order to reconcile the apparent contradiction posed by Fig. 5.4, however, we must come to a somewhat different conclusion concerning the power spectrum itself.

Suppose one considers pooling the spectral densities for one and the same heterojunction in (110) and (1–10) cross sections; these interface–resolved, cross–section averaged PSDs are presented side–by–side in Fig. 5.5. The power spectra in this case look indistinguishable from one another and, furthermore, the multi–component behavior first discerned in Fig. 5.3 is now emphasized by an improved signal–to–noise that follows with a larger, physically–equivalent ensemble.

Consider, next, an alternative, but not unreasonable, ensemble of equivalent size formed by combining the power spectral densities from both interface types – InAlAs–on–InGaAs with InGaAs–on–InAlAs – on one and the same cross section; these interface–averaged, cross–section resolved PDSs are compared in Fig. 5.6. The two spectra are, again, nearly indistinguishable from one another (indicating no strong dependence on cleavage direction) and, moreover, look no different than the pair of interface–resolved spectra in Fig. 5.5.

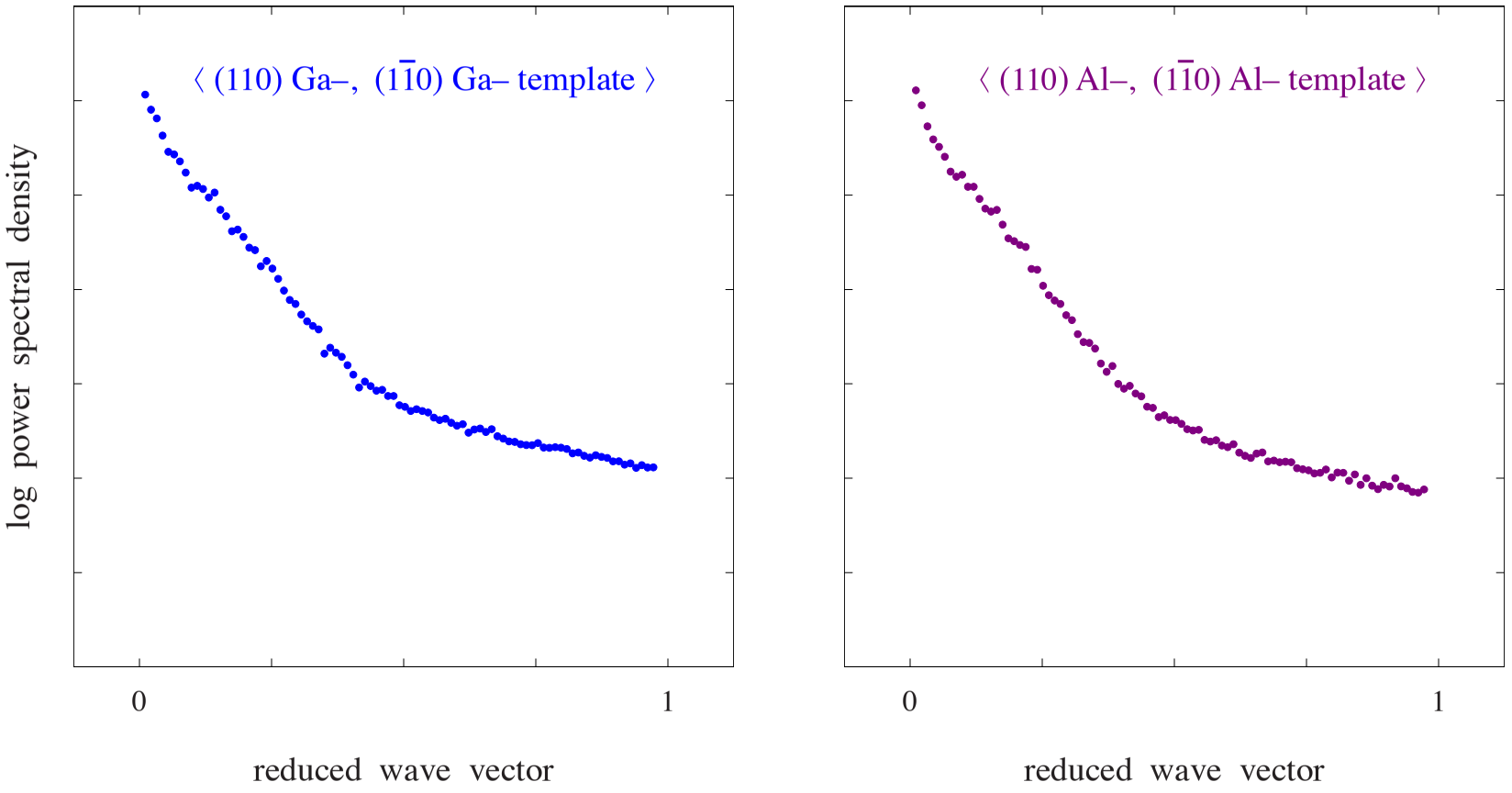


FIGURE 5.5. Cross–section–averaged, growth–template–resolved power spectral densities from pooled, –2.50 V data.

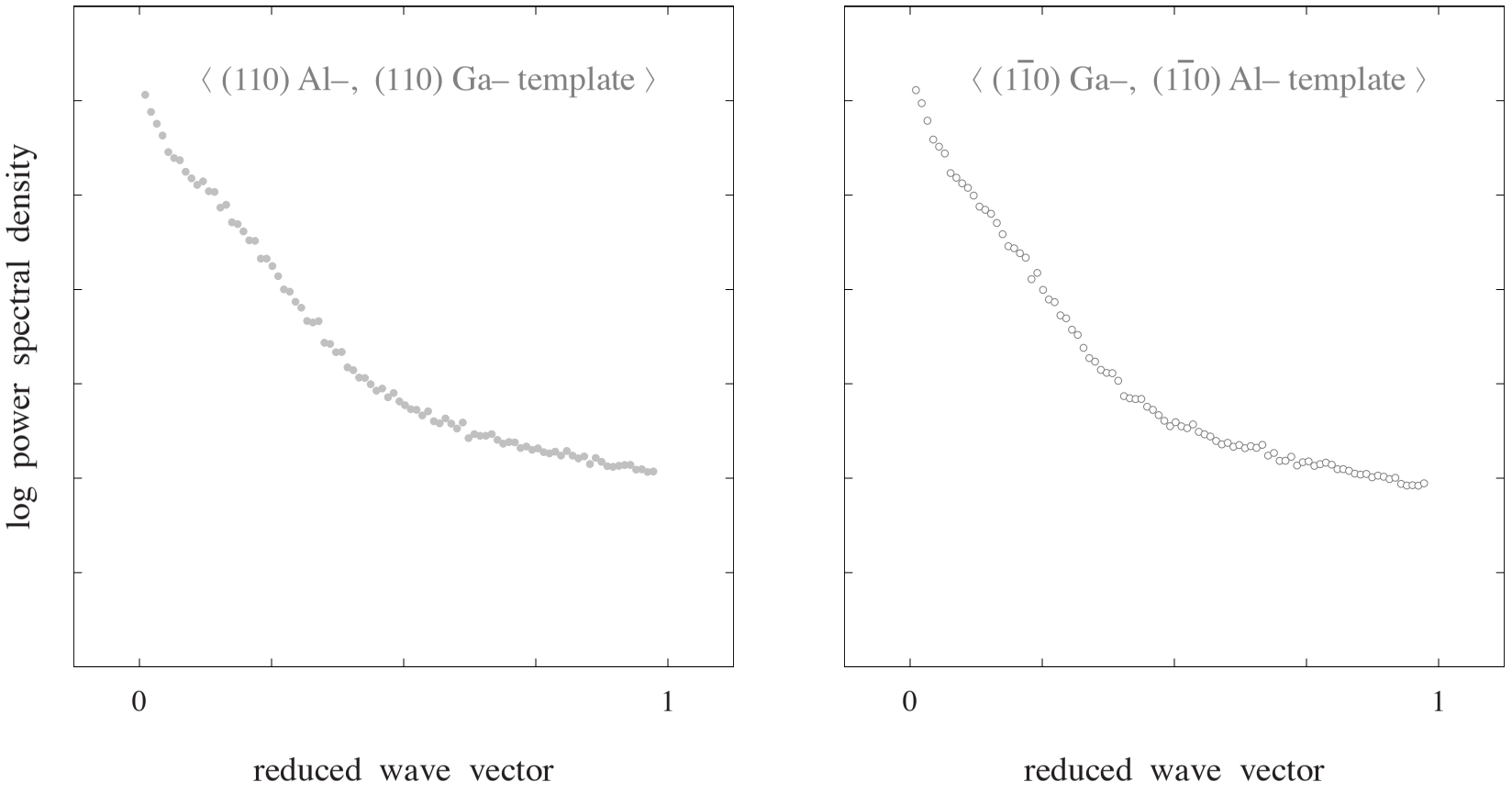


FIGURE 5.6. Cross–section–resolved, growth–template–averaged power spectral densities from pooled, –2.50 V data.

One further ensemble – purely combinatoric in nature and unrelated to any *a priori* meaningful physical criteria – may be formed by, again, pooling the two interface types as above, but this time from orthogonal cross–sections (i.e., combining InGaAs–InAlAs (110) with InAlAs–InGaAs (1–10), and vice versa), as shown in Fig. 5.7. Once again, these power spectra are uncannily similar, and nearly indistinguishable from the previous pairings shown in Figs. 5.5 and 5.6[2].

The lesson drawn from this exercise is the following: no matter which pair of original–ensemble PSDs selected for pooling, the outcome is the same: exponential behavior at small wave vectors ($q \lesssim 0.05$), together with a Gaussian–like component at intermediate ones ($0.05 \lesssim q \lesssim 0.30$). The consistency in the spectral density obtained by doubling the number of independent segments per "ensemble" suggests each of our original ensembles is physically indistinguishable from all others, and that any apparent differences in the behavior of their respective power spectral densities is reasonably attributed to limited statistics.

The logic just outlined justifies our subsequent pooling of all four, –2.50 V power spectral densities to obtain a representative picture of the spatial frequency distribution of interface roughness in this sample. The resulting "grand ensemble" power spectral density is shown in Fig. 5.8, where the multi–component behavior described above is now unmistakable.

---

[2] One may question why the PSDs in Fig. 5.2, as well as those in Fig. 5.3, might not also be so combined; the reason is the respective surveys at –2.50 V and –2.25 V were taken over the same area, i.e., with nearly identical ensembles of segments. That this is the case is clear once the PSDs for the two voltages are combined; the resulting signal–to–nose is closer to that of the original, unpooled data than the "double ensembles" assembled in Figs. 5.5 –5.7.

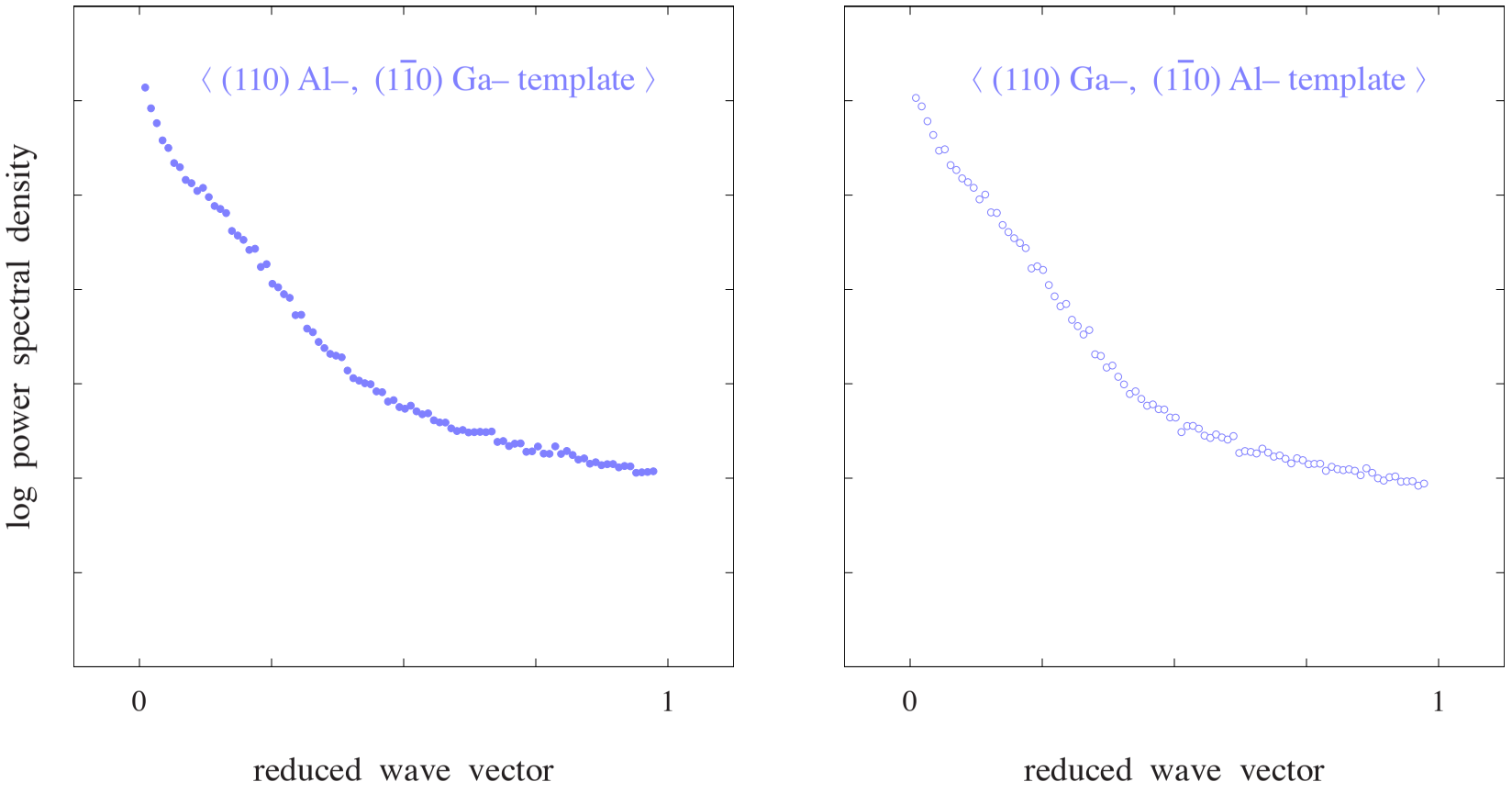


FIGURE 5.7. Mixed cross–section and growth–template poolings of –2.50 V data, unrelated to any *a priori* physical criteria.

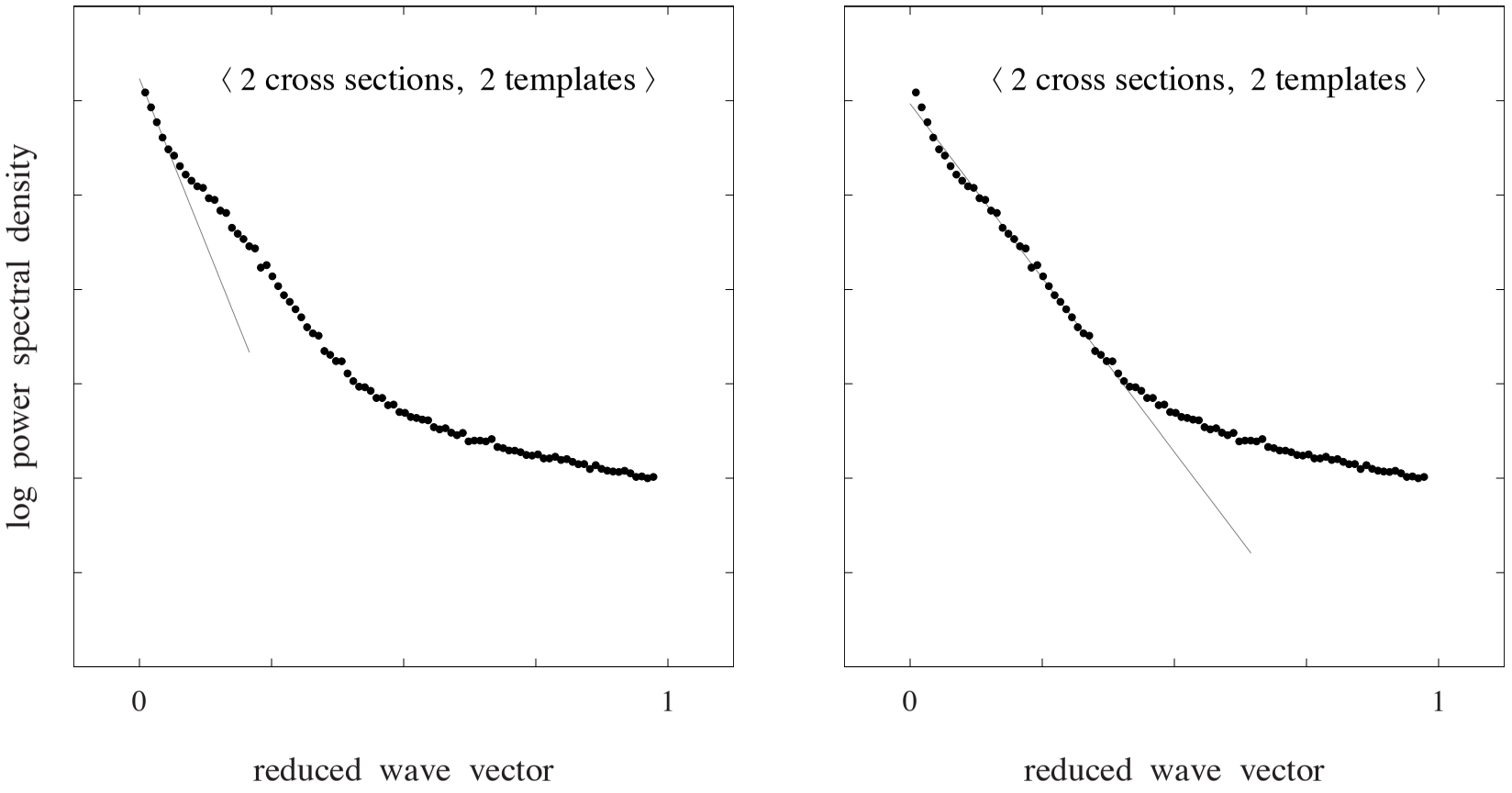


FIGURE 5.8. Cross–section averaged, growth–order averaged, power spectral density from pooling all –2.50 V data. This grand–ensemble spectrum is fit to a decaying exponential (grey) over the (spatial) frequency range $0 \leq q \leq \sim 0.05$ (left), and $0 \leq q \leq \sim 0.30$ (right).

Absent a sophisticated fitting algorithm that can properly account for the observed complexity of this spectral density, we fit the data only to an exponential–decay model within two restricted frequency ranges: ($0 < q \lesssim 0.05$) and ($0 < q \lesssim 0.30$), where the respective curves (grey) are shown in Fig. 5.8 (left) and (right); these fits are directly compared on a linear scale in Fig. 5.9, where differences between the two may be better appreciated. The real–space correlation lengths calculated from the corresponding fit parameters via (5.11) are summarized in Table 5.2. We note good agreement between the "grand ensemble" global exponential description in Table 5.2, and the original ensemble values in Table 5.1, consistent with the physical "indistinguishability" hypotheses underlying our "grand ensemble" pooling; as before, all quoted errors are from the associated least–squares fit and therefore likely to be too small.

Unanswered Questions

The analysis just described raises a number of important questions, many of which are not easily addressed. One obvious concern regarding the “grand–ensemble” pooling adopted here is the evidence in support of it, though logically consistent, is indirect; it would surely be more re–assuring if the number of interface segments belonging to each original–ensemble classification were sufficient to independently confirm the universality concluded above; whether this may yet be accomplished with the existing STM data is an ongoing discussion.

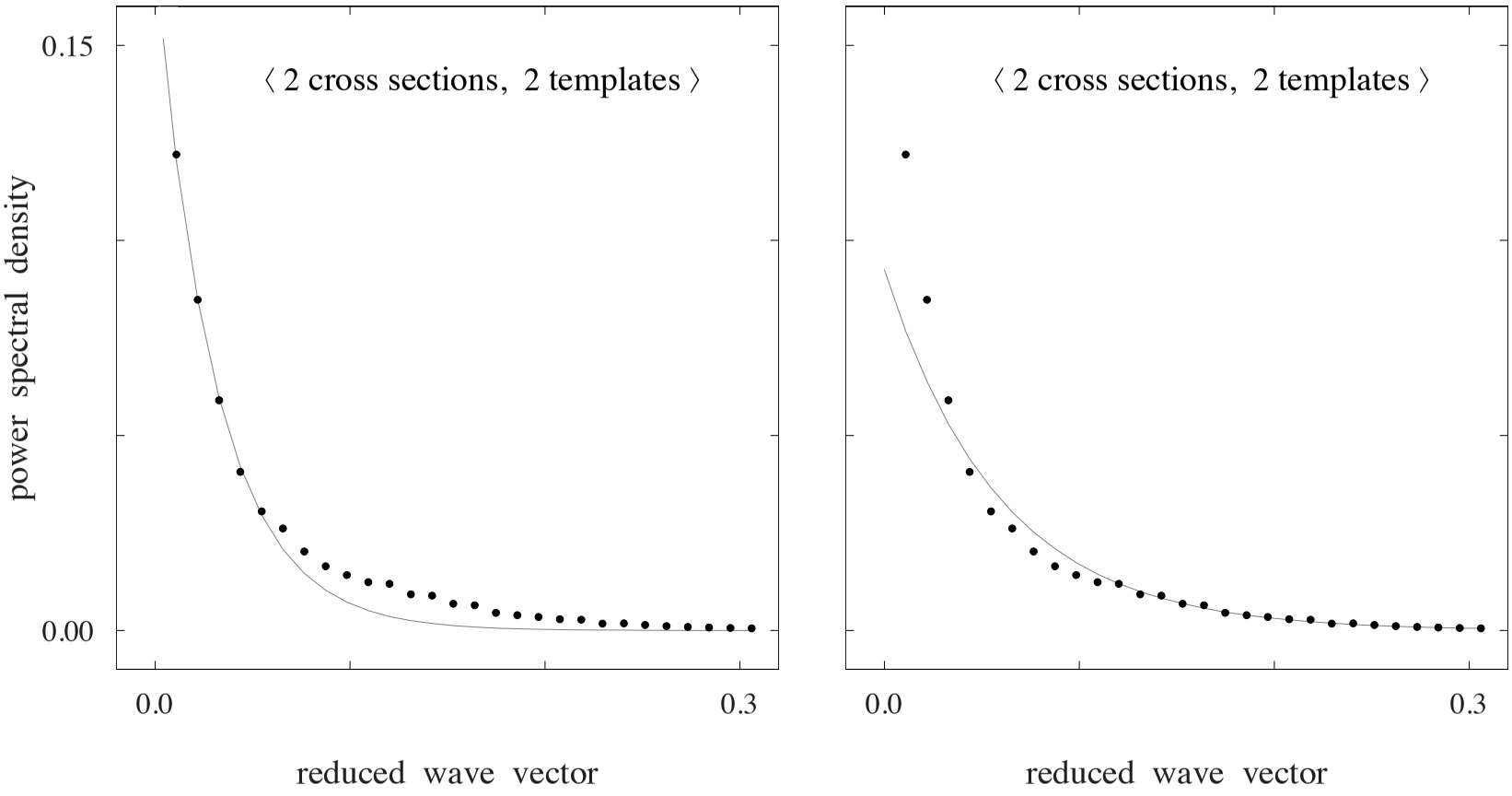


FIGURE 5.9. Grand ensemble power spectrum and exponential fits in Fig. 5.8 re–plotted on a linear (versus logarithmic) scale.

Table 5.2. Correlation lengths (Λ) obtained from exponential fits to the period–, cross–section, and growth–template pooled spectrum in Fig. 5.8. Quoted uncertainties are, again, uncorrelated, least–squares fitting errors and therefore likely too small.

$0 <$ reduced wave vector $\lesssim 0.05$		$0 <$ reduced wave vector $\lesssim 0.30$	
Λ	segments	Λ	segments
$(\mathbf{a}_{110})$	(M)	$(\mathbf{a}_{110})$	(M)
5.12 ± 0.13	153	**2.71 ± 0.05**	153

A likewise–obvious (and not unrelated) criticism is that further work is needed to develop a self–consistent fitting algorithm which correctly accounts for the multi–component, functional form of the observed power spectral density over the physically–relevant, small–wave–vector regime, as well as the high–frequency noise introduced by (device–irrelevant) sampling and digitization.

Additional work is also needed to understand the influence on these power spectra of the Fermi filter relied upon in Chapter III to remove high–spatial–frequencies (atomic corrugation) from the STM images and delineate interface profiles. Fig. 5.10 illustrates the power spectrum of this filter in relation to the grand–ensemble fits in Fig. 5.8, though it is by no means clear such comparisons are even meaningful. What is needed, instead, is a systematic evaluation of the dependence of our end–point power spectra on Fermi–filter parameters; that is a well–defined, but time–consuming, task.

Finally, the wholly–unexpected finding that our roughness power spectra are predominantly exponential rather than Gaussian begs for cogent physical explanation. What underlying physical process accounts for this functional form as well as its universality? Is the (apparent) irrelevance of physically–distinguishing factors such as growth order and growth–plane anisotropy peculiar to MOCVD or representative of MBE as well? Are the relative weights of exponential versus Gaussian components in the power spectrum, or their respective correlation lengths, predictable and / or controllable in any way?

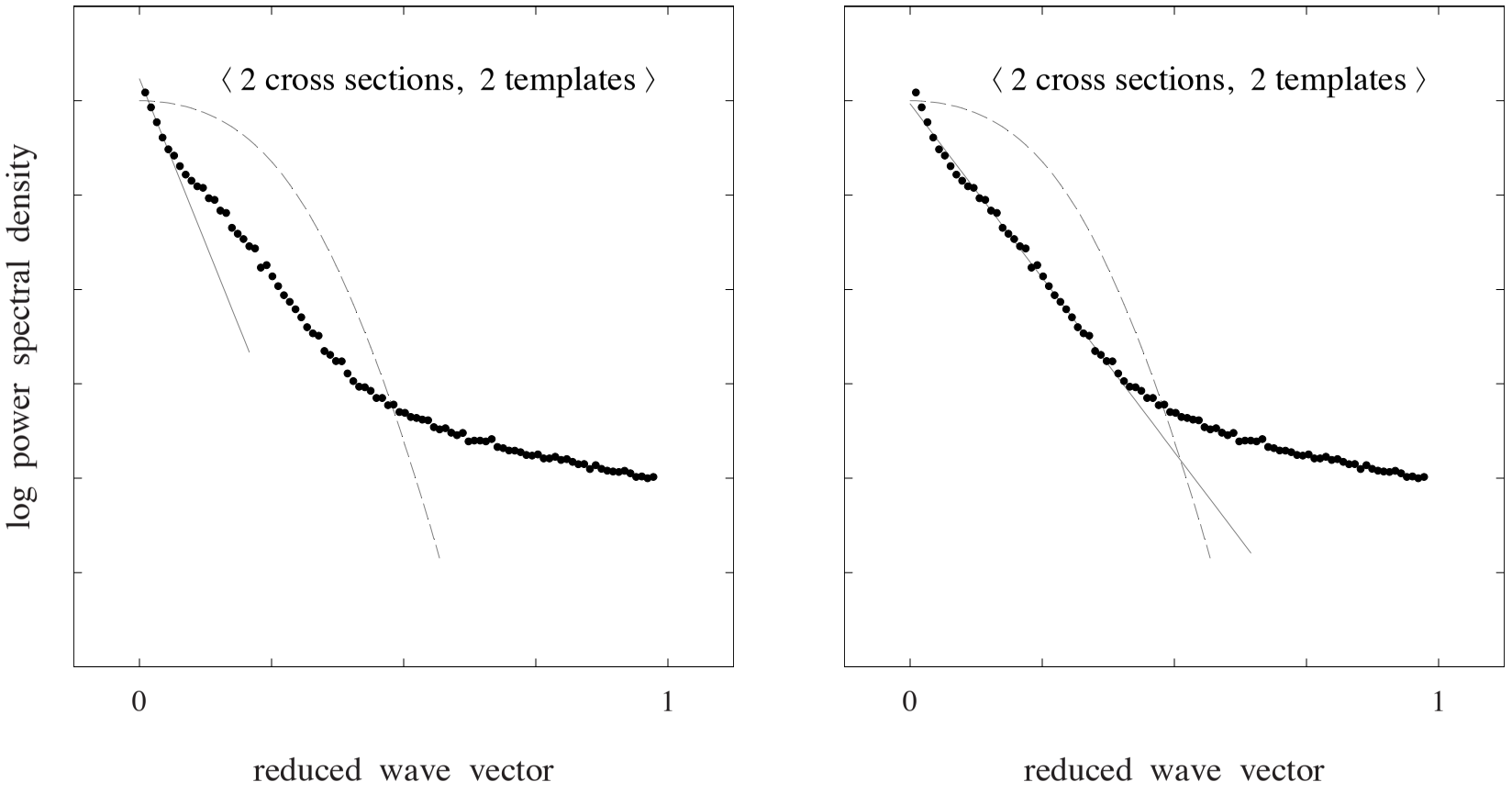


FIGURE 5.10. Grand–ensemble power spectrum (and exponential fits) in Fig. 5.8 compared with the power spectrum of the Fermi filter (dashed) used to remove high–frequency, atomic corrugation prior to STM–image thresholding (c.f., Figs. 3.5 – 3.8).

CHAPTER V

CONCLUSIONS

We have employed cross–sectional scanning tunneling microscopy to obtain an accurate statistical representation of the *as–grown* heterojunctions delineating the quantum wells and quantum barriers in InAlAs / InGaAs strain–balanced superlattices grown by MOCVD intimately related to quantum cascade lasers. Small deviations from the presumed planarity of these interfaces are believed to have profound implications for device performance, with energy–level broadening and carrier scattering as immediate, unintended consequences.

We have shown how the aluminum–rich barrier layers in these structures present significant challenges for present–day vacuum technology, and described the monumental efforts required to create and maintain a suitably–pristine habitat where freshly–exposed, aluminum–rich surfaces will remain clean over the many days needed to conclude representative STM surveys.

We described the development of carefully–constructed navigation protocols minimizing the image distortion inherent to STM piezo–electric raster mechanisms. These advances were used to implement a novel, reciprocal–space technique (analogous to Bragg's law in x–ray diffraction) to obtain local distance metrics insensitive to STM raster non–idealities. The method's accuracy was demonstrated with local period measurements that agree (to within hundredths of monolayers) with high–resolution x–

ray diffraction. With this new method we were able to establish a small, but nevertheless, measurable inconsistency between the targeted and *as–grown* superlattice periodicity on the order of 4 %.

We developed robust image processing algorithms that incorporate statistical criteria to reproducibly identify the interfaces separating quantum wells from barriers in cleavage–exposed cross–section. The heterojunction profiles obtained this way provide an experimentally–accessible avenue for delineating the confinement potential's spatial boundaries that appears logically consistent and physically reasonable.

We conducted a systematic analysis of interface roughness from the viewpoint of fluctuations about an experimentally–determined profile mean. Subtleties central to a correct understanding of roughness variance and its uncertainty in the face of correlated fluctuations are addressed, and the ensuing discussion illustrated with, and corroborated by, numerical simulations. Heterojunction growth order and growth–plane anisotropy are conclusively established as distinguishing physical characteristics. Our results portray a worrisome picture concerning the futility of fabricating extremely thin barriers (~ 4 ML, c.f., Fig. 1.6) in QCL devices since one can expect interface fluctuations on the order of ± 2.3 ML roughly 95 % of the time assuming adjacent interfaces are uncorrelated, but thickness fluctuations exceeding this amount will occur at least 5% of the time.

Finally, the power spectra of these interface fluctuations were analyzed to ascertain the correlation lengths and functional forms that govern their respective spatial–frequency dependencies. The available data are consistent with a universal power spectrum for MOCVD roughness that is isotropic, independent of heterojunction growth

order, and predominantly exponential in nature. The presence of multiple length scales in the power spectrum requires at least two, disparate physical origins for its explanation: atomic–scale alloy order, with correlation lengths of 1–2 lattice constants, and indium clustering, with correlations over 5–6 lattice constants, suggest themselves as reasonable physical mechanisms broadly consistent with our observations.

REFERENCES

[1] R. F. Kazarinov and R. A. Sirus, Soviet Physics Semiconductors **5**, 707 (1971).

[2] L. Esaki and R. Tsu, IBM Journal of Research and Development **14**, 61 (1970).

[3] R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys, and R. O. Carlson, Physical Review Letters **9**, 366 (1962).

[4] I. Hayashi, Appl. Phys. Lett. **17**, 109 (1970).

[5] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, Science **264**, 553 (1994).

[6] A. Rogalski, Progress in Quantum Electronics **27**, 59 (2003).

[7] C. Sirtori, H. Page, and C. Becker, Philosophical Transactions: Mathematical, Physical and Engineering Sciences **359**, 505 (2001).

[8] J. Devenson, D. Barate, O. Cathabard, R. Teissier, and A. N. Baranov, Applied Physics Letters **89**, 191115 (2006).

[9] Y. Chiu, Y. Dikmelik, P. Q. Liu, N. L. Aung, J. B. Khurgin, and C. F. Gmachl, Applied Physics Letters **101**, 171117 (2012).

[10] N. Zettili, *Quantum Mechanics: Concepts and Applications* (John-Wiley, Chichester, UK, 2001).

[11] G. E. P. Box, *Time Series Analysis Forecasting and Control* (Holden–Day, San Francisco, CA, 1976).

[12] H. Sakaki, T. Noda, K. Hirakawa, M. Tanaka, and T. Matsusue, Applied Physics Letters **51**, 1934 (1987).

[13] Y. C. Cheng and E. A. Sullivan, Surface Science **34**, 717 (1973).

[14] S. M. Goodnick, R. G. Gann, D. K. Ferry, C. W. Wilmsen, and O. L. Krivanek, Surface Science **113**, 233 (1982).

[15] M. A. Herman and H. Sitter, *Molecular Beam Epitaxy : Fundamentals and Current Status* (Springer - Verlag, Heidelberg, DE, 1996).

[16] M. Razeghi, *The MOCVD Challenge: a Survey of GaInAsP-InP and GaInAsP-GaAs for Photonic and Electronic Device Applications* (Taylor & Francis, Boca Raton, FL, 2010).

[17] G. B. Stringfellow, *Organometallic Vapor–Phase Epitaxy: Theory and Practice* (Academic Press, San Diego, CA, 1999).

[18] D. E. Aspnes and A. A. Studna, Physical Review Letters **54**, 1956 (1985).

[19] E. Colas, D. E. Aspnes, R. Bhat, A. A. Studna, M. A. Koza, and V. G. Keramidas, Journal of Crystal Growth **94**, 613 (1989).

[20] J. Steinshnider, J. Harper, M. Weimer, C. H. Lin, S. S. Pei, and D. H. Chow, Physical Review Letters **85**, 4562 (2000).

[21] J. Steinshnider, M. Weimer, R. Kaspi, and G. W. Turner, Physical Review Letters **85**, 2953 (2000).

[22] H. W. M. Salemink and O. Albrektsen, Physical Review B **47**, 16044 (1993).

[23] J. Harper, M. Weimer, D. Zhang, C. H. Lin, and S. S. Pei, Applied Physics Letters **73** (1998).

[24] P. Offermans, Appl. Phys. Lett. **82**, 1191 (2003).

[25] P. Offermans, Appl. Phys. Lett. **83**, 4131 (2003).

[26] N. Yao (private communication).

[27] J. A. Stroscio and W. J. Kaiser, *Scanning Tunneling Microscopy* (Academic Press, Boston, MA, 1993).

[28] C. B. Duke, Journal of Vacuum Science & Technology A **10**, 2032 (1992).

[29] R. M. Feenstra, J. A. Stroscio, J. Tersoff, and A. P. Fein, Physical Review Letters **58**, 1192 (1987).

[30] J. Tersoff and D. R. Hamann, Physical Review B **31**, 805 (1985).

[31] R. M. Feenstra and A. P. Fein, Physical Review B **32**, 1394 (1985).

[32] J. D. Steinshnider, Ph. D. Dissertation, Texas A&M University, 2002.

[33] M. R. Wood, K. Kanedy, F. Lopez, M. Weimer, J. F. Klem, S. D. Hawkins, E. A. Shaner, and J. K. Kim, Journal of Crystal Growth (2015).

[34] Z. Liu, C. F. Gmachl, C. G. Caneau, and C. Zah, Conference on Lasers and Electro-Optics (CLEO), 2 (2008).

[35] Omicron NanoTechnology GmbH., (McMurray, PA).

[36] G. A. Lengel, Ph. D. Dissertation, Texas A&M University, 1995.

[37] G. W. Brown, Ph. D. Dissertation, Texas A&M University, 1997.

[38] J. Harper, M. Weimer, D. Zhang, C. H. Lin, and S. S. Pei, in *Compound Semiconductors, 1997 IEEE International Symposium on*, 1998), p. 259.

[39] J. Harper, M. Weimer, D. Zhang, C. H. Lin, and S. S. Pei, Applied Physics Letters **73**, 2805 (1998).

[40] J. Harper, M. Weimer, D. Zhang, C. H. Lin, and S. S. Pei, (AVS, Ann Arbor, Michigan (USA), 1998), p. 1389.

[41] M. Zhong, J. Steinshnider, M. Weimer, and R. Kaspi, Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures **22**, 1593 (2004).

[42] MDC Vacuum Products LLC., (Hayward, CA).

[43] Gulf Coast Data Concepts LLC., (Waveland, MS).

[44] E. F. Schubert, (Cambridge Univ. Press, 2015), Vol. 2015.

[45] F. R.M, Physica B: Condensed Matter **273-274**, 796 (1999).

[46] J. H. Davies, P. Offermans, and P. M. Koenraad, Journal of Applied Physics **98**, 053504 (2005).

[47] E. P. Stoll, Ultramicroscopy **42-44, Part 2**, 1585 (1992).

[48] J. F. Jorgensen, K. Carneiro, L. L. Madsen, and K. Conradsen, (AVS, Beijing, China, 1994), p. 1702.

[49] M. Hunstig, T. Hemsel, and W. Sextro, Sensors and Actuators A: Physical **200**, 79 (2013).

[50] P. F. Fewster, *X-ray Scattering from Semiconductors* (Imperial College Press, London, UK, 2003).

[51] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, Boston, MA, 2003).

[52] R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (Prentice Hall, Upper Saddle River, NJ, 2008).

[53] J. D. Murray and W. VanRyper, *Encyclopedia of Graphics File Formats* (O'Reilly & Associates, 1996).

[54] Adobe Systems Inc., (San Jose, CA).

[55] A. Zangwill, *Physics at Surfaces* (Cambridge University Press, Cambridge, UK, 1988).

[56] J. Y. Tsao, *Materials Fundamentals of Molecular Beam Epitaxy* (Academic Press, San Diego, CA, 1993).

[57] M. W. Wang, D. A. Collins, T. C. McGill, R. W. Grant, and R. M. Feenstra, Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures **13**, 1689 (1995).

[58] M. B. Priestley, *Spectral Analysis and Time Series* (Academic Press, London, UK, 1981).

APPENDIX A

TECHNICAL NOTES

Image Processing and Analysis Pipeline

We present in Table A.1 the list of image processing steps (*i* to *xiii*) employed to digitally identify and extract heterojunction profiles from STM images as described in Chapter III. Also in Table A.1, are the analytical steps (*xiv* and *xv*) used for the subsequent statistical evaluation of the interface roughness variance and corresponding spectral frequency distribution as detailed in Chapters IV and V, respectively.

STM images are resampled in step *vi* from the native 100 pixels–per–inch (ppi) resolution to 400 ppi to increase the angular resolution necessary to determine interface angles. These angles – corrected by substrate vicinality (*ix*) – are then used to rotate "low–pass" filtered images (which are subsequently interpolated and resampled) in step *vii* so that every increment of the heterojunction profiles in the [001] growth direction as well as in the [110] in–plane direction is equally sampled – a requirement necessary to compute a valid power spectrum.

The visual effect that image resolution has on the experimentally defined heterojunctions can be seen in Figure A.1 where two interfaces sampled at the native STM resolution of 100 ppi, left, are resampled to 400 ppi, right. It is clear from this comparison that the heterojunction profiles appear smoother at higher resolutions, but this effect has almost no bearing on the roughness variance (Chapter IV). It does,

however, influence the roughness spectrum (Chapter V), where the amplitude of high–spatial–frequency digitization noise is suppressed (relative to the original 1Å per pixel digitization) and “brown” noise from finite–length sampling clearly identified.

Table A.1. Image processing and analysis steps in the characterization of interface roughness and correlations as described in Chapters III–V.

i	use of 750 Å in *x* and *y* images
ii	amplify grey–level intensity by a factor of two
iii	crop image to 700 Å in *x* and *y*
iv	apply second order background subtraction
v	crop image to 700 Å in *x* and 350 Å *y*
vi	resample cropped image to determine heterojunction angles
vii	apply (strong) Fermi filter to step (*v*)
viii	determine Otsu threshold
ix	incorporate vicinality correction to heterojunction angles in (*vi*)
x	rotate, interpolate, and resample step (*vii*)
xi	replace grey levels below or above threshold with black or white
xii	crop individual segments using a fixed length
xiii	extract heterojunction profile points using Laplacian
xiv	form an ensemble pool of vertical fluctuations normalized to the survey averaged [001] lattice constant
xv	form an ensemble pool of power spectra whose independent variable – the spatial frequency – is normalized to the survey average [110] reciprocal lattice vector

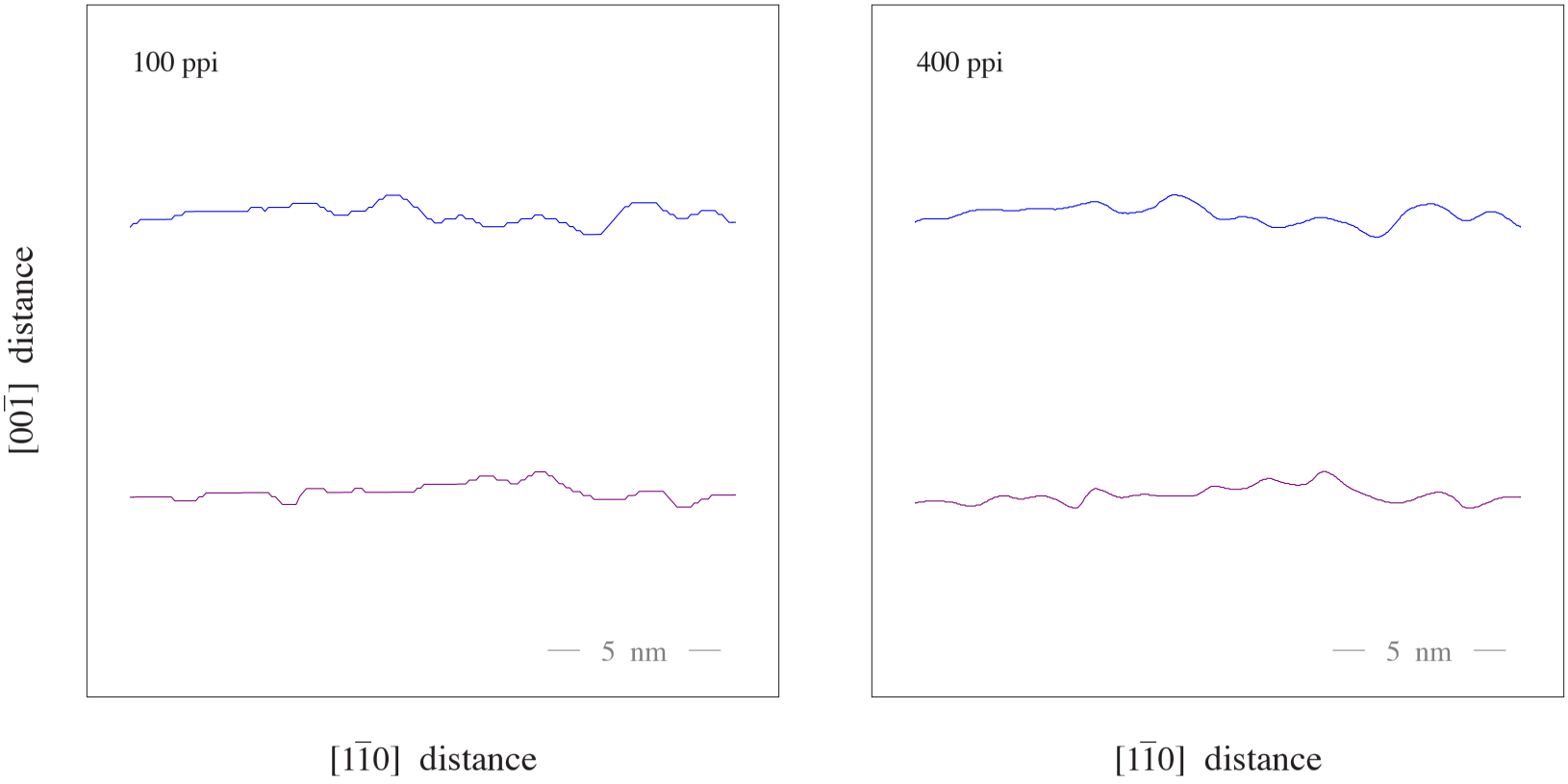


FIGURE A.1. Comparison of digitized interface profiles with native (100 ppi) and resampled (400 ppi) resolutions. Higher–resolution profiles provide a closer approximation to an ideally–continuous stochastic process.